

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.

SEARCH REQUEST FORM

3-423

Requestor's Name: T. Cunningham Serial Number: 08/340,510
Date: 3/23 Phone: 3968 Art Unit: 1813

Search Topic:

Please write a detailed statement of search topic. Describe specifically as possible the subject matter to be searched. Define any terms that may have a special meaning. Give examples or relevant citations, authors keywords, etc., if known. For sequences, please attach a copy of the sequence. You may include a copy of the broadest and/or most relevant claim(s).

Search AA d.b.'s for best
match for:

SEQ ID's 1 and 2

STAFF USE ONLY

Date completed: 3-24-95
Searcher: ABR
Terminal time: _____
Elapsed time: _____
CPU time: _____
Total time: _____
Number of Searches: _____
Number of Databases: _____

Search Site
____ STIC
____ ☒ CM-1
____ Pre-S
Type of Search
____ N.A. Sequence
____ A.A. Sequence
____ Structure
____ Bibliographic

Vendors
____ ☒ IG Suite
____ ☒ STN
____ Dialog
____ APS
____ Geninfo
____ SDC
____ DARC/Questel
____ Other

Dep Co.
Q 1210
7a

EVALUATION OF MPSRCH - A NEW TOOL FOR SEQUENCE SEARCHING.

STIC is in the process of evaluating the MPSRCH software for conducting your sequence searches on our new MASPAR (massively parallel processing computer). This software was written by John Collins and S. Sturrock at Edinburgh University, U.K. and is distributed by Intelligenetics, Inc.

Some of the advantages of the MPSRCH /MASPAR combination over our currently used configuration of FASTDB/SUN workstations, are as follows:

1. Speed: The MASPAR machine has 16,000 processors as compared to 8 on our fastest SUN server, and is therefore capable of millions of cell/updates/sec. For example, a protein search taking 30 minutes on Sun was completed in 25 seconds on MASPAR.
2. The ability to run the full Smith-Waterman optimization algorithm. This is especially important with DNA searches as we cannot currently justify doing this on Sun due to the substantially increased runtimes. The DNA search results that you presently receive from Fastdb are for the initial score only, in most cases this alignment is satisfactory, however the optimization step offered in MPSRCH ensures the best ranking of the initial scores so that the "better" alignments may be located closer to the top of the alignment table.
3. Complementary (inverse) DNA strand is automatically searched in MPSRCH. These results are denoted by a small letter "c" adjacent to the Result No. in the SUMMARIES table.
4. Full support for all the IUPAC Nucleic Acid Codes Ambiguity codes when appropriate.

MPSRCH/MASPAR format vs. FASTDB/SUN format

1. Lack of display context:
At the present time there is no user selectable display context available, thus you cannot see 10-50 bases on either side of the aligned sequence, as is available in FASTDB/SUN format.

2. Match % anomaly: (See Example 1 below).
Look at Result No. 7 in the SUMMARIES table and in the alignment.

Note :

a: That in the SUMMARIES table the % Query Match is defined as the percentage of the query sequence matched (the score of the search, 18 in this example), calculated as a percentage of the perfect score found in the alignment (17 in this example) . It is determined as follows:

1. The perfect score is calculated as the score of the query against itself (18 in this example)
2. The score of the query against database is determined (17 in this example)
3. The % query match is calculated as:

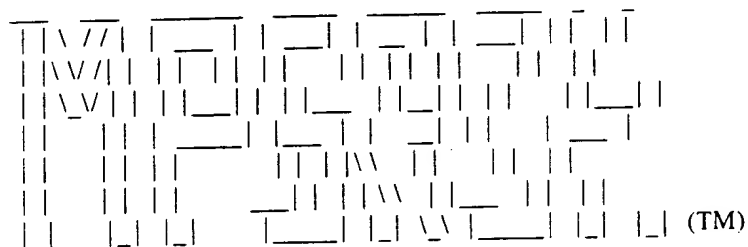
$$\frac{\text{score}}{\text{perfect score}} * 100 \quad \text{in this example} \quad \frac{17}{18} * 100 = 94.4 \%$$

b. The match % shown in the annotation is called the "Best Locally Similar Alignment " . This is defined as the point where no further improvement in the score can be obtained, even if the search is continued to the ends of the sequences. The aligned region shown thus represents the best possible continuation of the alignment. In this example, it is calculated thus:

$$\frac{\text{score}}{\text{Matches}} * 100 \quad \text{in this example} \quad \frac{17}{17} * 100 = 100.0 \%$$

The Examiners are cautioned to inspect both the % Query Match in the SUMMARIES table and Match % in the annotation in view of the above explanation, before using any alignment for rejection. Future versions of the MPSRCH software will incorporate both % match values in the annotation.

Example 1.



Release 2.0 John F. Collins & S. S. Sturrock, Biocomputing Research Unit.

Copyright (c) 1993, 1994 by University of Edinburgh, U.K.

Distribution rights by IntelliGenetics, Inc.

MPsrch_nn n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Thu Feb 23 09:09:16 1995; MasPar time 4.49 Seconds

195.126 Million cell updates/sec

Tabular output not generated.

Title: >US-08-XXX-XXX-X

Description: (1:18) from US08XXXXXX.seq

Perfect Score: 18

N.A. Sequence: TTAGGGTTAGGGTTAGGG 18

Comp: AATCCCAATCCCAATCCC

Gap 60Nmatch STD : Dbase 0; Query 0

Searched: 57621 seqs, 24347505 bases x 2

Database: n-geneseq

Statistics: Mean 3.876; Variance 2.020; scale 1.919

Predicted No. is the number of results expected by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result			% Query		DB ID	Description	Pred. No.
	No.	Score	Match	Length			
c 1	18	100.0	24	10	Q52411	Human telomere length	3.40e-04
c 2	18	100.0	20	10	Q51150	Human telomeric singl	3.40e-04
3	18	100.0	60	10	Q51147	Human telomeric singl	3.40e-04
4	18	100.0	24	10	Q52413	Human CD4+ lymphocyte	3.40e-04
5	18	100.0	36	11	Q63638	Human herpes virus 7	3.40e-04
6	18	100.0	18	10	Q52410	Human telomere length	3.40e-04
c 7	17	94.4	163	1	Q06654	Feline T-cell lymphot	1.62e-03

ALIGNMENTS

RESULT 7

ID Q06654 standard; DNA; 163 BP.AC Q06654;
 DT 26-FEB-1991 (first entry)
 DE Feline T-cell lymphotropic lentivirus of clone 2BYCXL2.
 KW Feline T-cell lymphotropic lentivirus; FIV; 2BYCXL2; antibodies;
 KW vaccines; ds.
 OS Feline T-cell lymphotropic lentivirus 2428 (Pentaluma).
 FT CDS 2..163
 FT /*tag= a
 FT /label=FIV
 PN WO9013573-A.
 PD 15-NOV-1990.
 PF 30-APR-1990; U02338.
 PR 08-MAY-1989; US-348784.
 PR 08-DEC-1989; US-447810.
 PA (IDEX-) IDEXX CORP.
 PI Anderson PR, Oconnor TP, Tonelli QJ;
 DR WPI; 90-361429/48.
 DR P-PSDB; R08085.
 PT Feline T-cell lympho-tropic lentivirus poly-peptide(s) - used for
 PT specific detection of FIV antibodies, prodn. of antibodies and in
 PT vaccines
 PS Disclosure; Fig 5(b); 37pp; English.
 CC See also Q06653-55 and R08094-96.
 SQ Sequence 163 BP; 28 A; 66 C; 37 G; 30 T;

DB 1; Score 17; Match 100.0%; Predicted No. 1.62e-03;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 95 cctaaccctaaccctaa 111
 |||||
 Cp 17 CCTAACCCTAACCCTAA 1

SERIAL NUMBER:

DATE SEARCHED:

CONFIDENTIAL - FOR INTERNAL PTO USE ONLY.

EXAMINER INPUT REQUEST -EVALUATION OF SEQUENCE SEARCHES DONE ON THE MPSRCH/MASPAR SYSTEM.

Your sequence search request was run in parallel on the FASTDB/SUN system as well as the new MPSRCH/MASPAR configuration. STIC would really appreciate if you could take a few minutes of your time to provide the answers to the questions below. Your participation in this survey is completely optional. Please review the search results from the MPSRCH/MASPAR configuration in view of the FASTDB/SUN results and provide us feedback on the following questions:

1. What are your general impressions of the MPSRCH format? What did you like? What were your dislikes?

2. Did MPSRCH/MASPAR find any especially relevant alignments that were missed by the FASTDB/SUN search?

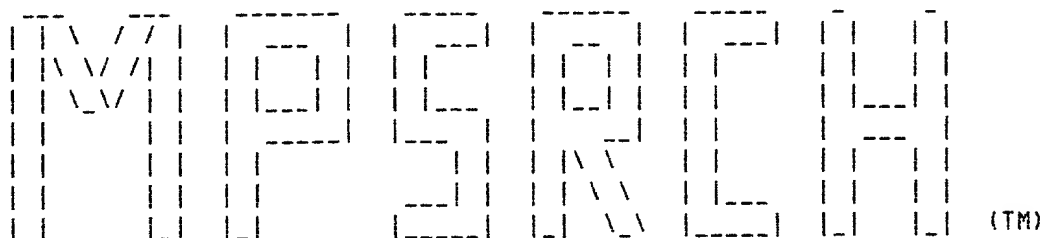
3. Did MPSRCH/MASPAR miss any especially relevant alignments that were found by FASTDB/SUN search?

4. Did MPSRCH/MASPAR find any especially relevant alignments in the top ten positions that were much lower down in the FASTDB/SUN alignment table?

5. Did you find the problem with the MPSRCH/MASPAR % alignment ambiguity (as illustrated by Example 1 above), and the lack of display context to be a major hindrance in your understanding of the MPSRCH/MASPAR results?

6. Any other comments? Thank You.

PLEASE RETURN THIS FROM TO SEARCH INPUT TRAY IN THE 12TH FLOOR COMPUTER CLUSTER OR TO THE REFERENCE DESK/INPUT TRAY IN THE BIOTECH/CHEM LIBRARY.



Release 2.0 John F. Collins & S. S. Sturrock, Biocomputing Research Unit.
Copyright (c) 1993, 1994 by University of Edinburgh, U.K.
Distribution rights by IntelliGenetics, Inc.

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Mar 24 07:42:16 1995; MasPar time 2.88 Seconds
56.789 Million cell updates/sec

Tabular output not generated.

Title: >US-08-300-510-1
Description: (1:27) from US08300510.pep
Perfect Score: 195
Sequence: 1 KRVDLFLTGPDEYVEQVAQYKALPV 27

Scoring table: PAM 150
Gap 14

Searched: 50375 seqs, 6065180 residues

Database: a-geneseq
1 a-gen1
2 a-gen2
3 a-gen3
4 a-gen4
5 a-gen5
6 a-gen6
7 a-gen7
8 a-gen8
9 a-gen9
10 a-gen10

Statistics: Mean 22.071; Variance 75.122; scale 0.294

Predicted No. is the number of results expected by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description	Pred. No.
1	195	100.0	88	8 R41984	Human T cell reactive	1.14e-13
2	195	100.0	92	8 R41983	Human T cell reactive	1.14e-13
3	195	100.0	27	8 R41975	Human T cell reactive	1.14e-13
4	195	100.0	96	7 R36548	Recombitope YZX.	1.14e-13
5	195	100.0	94	3 R12119	TRFP chain 1 with lea	1.14e-13
6	195	100.0	92	7 R36539	TRFP chain 1 (with Le	1.14e-13
7	195	100.0	88	7 R36540	TRFP chain 1 (with Le	1.14e-13
8	195	100.0	27	7 R36542	Peptide X.	1.14e-13
9	195	100.0	96	3 R12120	TRFP chain 1 with lea	1.14e-13
10	195	100.0	94	5 R27749	TRFP chain 1 with lea	1.14e-13

11	195	100.0	94	5	R27367	TRFP Chain #1 with C1	1.14e-13
12	67	34.4	422	10	R54202	snaA gene product inv	8.62e+00
13	65	33.3	453	9	R47872	Enzyme/biocatalyst wh	1.33e+01
14	63	32.3	234	4	R20746	Human R-PTPase beta s	2.03e+01
15	63	32.3	452	3	R13119	Phenylalanine hydroxy	2.03e+01
16	63	32.3	704	1	P80087	Sequence of 85 kd pro	2.03e+01
17	63	32.3	235	4	R20748	Human R-PTPase gamma	2.03e+01
18	62	31.8	535	1	P90181	Cross-reactive materi	2.51e+01
19	59	30.3	389	2	P70668	D-alanine racemase.	4.71e+01
20	58	29.7	1480	3	R13300	CFTR Y563N.	5.79e+01
21	57	29.2	1480	3	R13297	CFTR S549R.	7.11e+01
22	57	29.2	383	10	R51059	Sequence of plasmid p	7.11e+01
23	57	29.2	1190	3	R13308	CFTR 3659 del C.	7.11e+01
24	57	29.2	1480	3	R13234	CFTR G178R.	7.11e+01
25	57	29.2	1100	1	P95644	Rabbit seletal muscle	7.11e+01
26	57	29.2	1480	4	R22492	Cystic Fibrosis trans	7.11e+01
27	57	29.2	1480	3	R13235	CFTR A455E.	7.11e+01
28	57	29.2	1480	3	R13894	Cystic fibrosis trans	7.11e+01
29	57	29.2	1480	3	R13299	CFTR R560T.	7.11e+01
30	57	29.2	1480	3	R13232	CFTR G85E.	7.11e+01
31	57	29.2	1480	2	R11115	Cystic fibrosis trans	7.11e+01
32	57	29.2	1479	2	R11602	Mutant cystic fibrosi	7.11e+01
33	57	29.2	1480	3	R13302	CFTR L1077P.	7.11e+01
34	57	29.2	1480	4	R23074	Cystic fibrosis gene	7.11e+01
35	57	29.2	1480	3	R13233	CFTR I148T.	7.11e+01
36	57	29.2	1479	3	R13231	CFTR delta I507.	7.11e+01
37	57	29.2	1091	3	R13303	CFTR Y1092X.	7.11e+01
38	57	29.2	1091	7	R33553	Sequence of the alpha	7.11e+01
39	57	29.2	1480	3	R13298	CFTR G551D.	7.11e+01
40	56	28.7	508	10	R51696	Human PDI.	8.72e+01
41	56	28.7	37	7	R38914	Recombitope ZXY.	8.72e+01
42	56	28.7	491	6	R25296	Recombinant PDI (Asp1	8.72e+01
43	56	28.7	508	6	R25297	PDI.	8.72e+01
44	56	28.7	515	10	R51697	Human HSA-PDI fusion	8.72e+01
45	56	28.7	508	1	P80664	Polypeptide with prot	8.72e+01

ALIGNMENTS

RESULT 1

ID R41984 standard; Protein; 88 AA.
AC R41984;
DT 21-APR-1994 (first entry)
DE Human T cell reactive feline protein B chain 1.
KW Human; T cell; reactive; feline; protein; immune response; antigen;
KW tolerance; mammal; Dermatophagoides; Felis; Ambrosia; Lolium; Canis;
KW Cryptomeria; Alternaria; Alder; Betula; Quercus; Olea; Artemesia;
KW Plantago; Parietaria; Blattella; Apis; Periplaneta; autoantigen.
OS Homo sapiens.
FH Key Location/Qualifiers
FT Peptide 1..17
FT /note= "Signal peptide"
FT Protein 18..88
FT /note= "Mature protein"
PN W09319178-A.
PD 30-SEP-1993.
PF 25-MAR-1993; U02462.
PR 25-MAR-1992; US-857311.
PR 15-MAY-1992; US-884718.
PR 15-JAN-1993; US-006116.
PA (IMMU-) IMMUNOLOGIC PHARM CORP.
PI Briner TJ, Garman RD, Gefter ML, Greenstein JL, Kuo M;
PI Morville M;
DR WPI; 93-320744/40.
DR N-PSDB; Q49534.
BT New amino acid sequence for induced release of gamma interferon

PT epitope(s) of an allergen administered subcutaneously, for
PT treating sensitivity to cats, bees, etc.
PS Disclosure: Fig 1; 107pp; English.
CC The sequences given in R41983-84 represent chain 1 of human T cell
CC reactive feline proteins (TRFP) A and B respectively. Peptides
CC derived from TRFP may be used in a therapeutic composition which is
CC useful in treating diseases which involve an immune response to a
CC protein antigen. This composition may be used to induce tolerance
CC in a mammal to Dermatophagoides, Felis, Ambrosia, Lolium, Cryptomeria,
CC Alternaria, Alder, Betula, Quercus, Olea, Artemesia, Plantago,
CC Parietaria, Canis, Blattella, Apis, Periplaneta and to autoantigens
CC in humans.
SQ Sequence 88 AA;

DB 8; Score 195; Match 100.0%; Predicted No. 1.14e-13;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 25 krdvdlfltgtptdeyveqvaqykalpv 51
|||||
Qy 1 KRVDLFLTGTPTDEYVEQVAQYKALPV 27

RESULT 2

ID R41983 standard; Protein; 92 AA.
AC R41983;
DT 21-APR-1994 (first entry)
DE Human T cell reactive feline protein A chain 1.
KW Human; T cell; reactive; feline; protein; immune response; antigen;
KW tolerance; mammal; Dermatophagoides; Felis; Ambrosia; Lolium; Canis;
KW Cryptomeria; Alternaria; Alder; Betula; Quercus; Olea; Artemesia;
KW Plantago; Parietaria; Blattella; Apis; Periplaneta; autoantigen.
OS Homo sapiens.
FH Key Location/Qualifiers
FT Peptide 1..22
FT /note= "Signal peptide"
FT Protein 23..92
FT /note= "Mature protein"
PN W09319178-A.
PD 30-SEP-1993.
PF 25-MAR-1993; U02462.
PR 25-MAR-1992; US-857311.
PR 15-MAY-1992; US-884718.
PR 15-JAN-1993; US-006116.
PA (IMMU-) IMMUNOLOGIC PHARM CORP.
PI Briner TJ, Garman RD, Geffer ML, Greenstein JL, Kuo M;
PI Morville M;
DR WPI; 93-320744/40.
DR N-PSDB; Q49533.
PT New peptide(s) for inducing tolerance - comprise one or more
PT epitope(s) of an allergen administered subcutaneously, for
PT treating sensitivity to cats, bees, etc.
PS Disclosure: Fig 1; 107pp; English.
CC The sequences given in R41983-84 represent chain 1 of human T cell
CC reactive feline proteins (TRFP) A and B respectively. Peptides
CC derived from TRFP may be used in a therapeutic composition which is
CC useful in treating diseases which involve an immune response to a
CC protein antigen. This composition may be used to induce tolerance
CC in a mammal to Dermatophagoides, Felis, Ambrosia, Lolium, Cryptomeria,
CC Alternaria, Alder, Betula, Quercus, Olea, Artemesia, Plantago,
CC Parietaria, Canis, Blattella, Apis, Periplaneta and to autoantigens
CC in humans.
SQ Sequence 92 AA;

DB 8; Score 195; Match 100.0%; Predicted No. 1.14e-13;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 29 krdvdlfltgtptdeyveqvaqykalpv 55
|||||
Qy 1 KRVDLFLTGTPTDEYVEQVAQYKALPV 27

RESULT 3

ID R41975 standard; peptide; 27 AA.
AC R41975;
DT 21-APR-1994 (first entry)
DE Human T cell reactive feline protein fragment X.
KW Human; T cell; reactive; feline; protein; immune response; antigen;
KW tolerance; mammal; Dermatophagoides; Felis; Ambrosia; Lolium; Canis;
KW Cryptomeria; Alternaria; Alder; Betula; Quercus; Olea; Artemesia;
KW Plantago; Parietaria; Blattella; Apis; Periplaneta; autoantigen; ss.
OS Homo sapiens.
PN W09319178-A.
PD 30-SEP-1993.
PF 25-MAR-1993; U02462.
PR 25-MAR-1992; US-857311.
PR 15-MAY-1992; US-884718.
PR 15-JAN-1993; US-006116.
PA (IMMU-) IMMUNOLOGIC PHARM CORP.
PI Briner TJ, Garman RD, Geffer ML, Greenstein JL;
PI Kuo M, Morville M;
DR WPI; 93-320744/40.
PT New peptide(s) for inducing tolerance - comprise one or more
PT epitope(s) of an allergen administered subcutaneously, for
PT treating sensitivity to cats, bees, etc.
PS Claim 1; Fig 3; 107pp; English.
CC The sequences given in R41975-82 are peptides derived from a human T
CC cell reactive feline protein. These peptides are used in a
CC therapeutic composition which is useful in treating diseases which
CC involve an immune response to a protein antigen. This composition
CC may be used to induce tolerance in a mammal to Dermatophagoides,
CC Felis, Ambrosia, Lolium, Cryptomeria, Alternaria, Alder, Betula,
CC Quercus, Olea, Artemesia, Plantago, Parietaria, Canis, Blattella,
CC Apis, Periplaneta and to autoantigens in humans.
SQ Sequence 27 AA;

DB 8; Score 195; Match 100.0%; Predicted No. 1.14e-13;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 krdvdlfltgtptdeyveqvaqykalpv 27
|||||
Qy 1 KRVDLFLTGTPTDEYVEQVAQYKALPV 27

RESULT 4

ID R36548 standard; Protein; 96 AA.
AC R36548;
DT 12-AUG-1993 (first entry)
DE Recombitope YZX.
KW Human T cell reactive feline protein; TRFP; epitope; recombiteope
KW sensitivity; Felis domesticus.
OS Synthetic.
FH Key Location/Qualifiers
FT Cleavage_site 14..15
FT /label= thrombin_cleavage_site
PN W09308280-A.
PD 29-APR-1993.
PF 16-OCT-1992; U08694.
PR 16-OCT-1991; US-777859.
PR 13-DEC-1991; US-807529.
PA (IMMU-) IMMULOGIC PHARM CORP.
PI Bond JF, Garman RD, Kuo M, Morgenstern JP, Morville M;
PI

DR WPI; 93-152473/18.
 DR N-PSDB; Q41572.
 PT Recombitope peptide having T-cell stimulating activity - for the
 PT diagnosis and treatment of sensitivity to protein allergens,
 PT auto:antigens and protein antigens
 PS Disclosure; Fig 8; 73pp; English.
 CC Preferred recombitepeptides for treating sensitivity to Felis
 CC domesticus are derived from the the genus Felis and comprise
 CC regions selected from peptides X, Y, Z, A and B, of TRFP, and
 CC modifications thereof, such as peptide C.
 CC Oligonucleotides C, D, E, F, G, H and I are used in the
 CC construction of recombitepeptide YZX.
 SQ Sequence 96 AA;

DB 7; Score 195; Match 100.0%; Predicted No. 1.14e-13;
 Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 70 krdvdlfltgtptdeyveqvaqykalpv 96
 |||||
 Qy 1 KRDVDLFLTGTPTDEYVEQVAQYKALPV 27

RESULT 5

ID R12119 standard; Protein; 94 AA.
 AC R12119;
 DT 26-JUL-1991 (first entry)
 DE TRFP chain 1 with leader A.
 KW Human T cell reactive feline protein; cat allergens.
 QS Felis catus.
 FH Key Location/Qualifiers
 FT Peptide 3..24
 FT /label= Leader B
 FT Protein 25..94
 FT /label= TRFP Chain 1
 PN WD9106571-A.
 PD 16-MAY-1991.
 PF 02-NOV-1990; U06548.
 PR 03-NOV-1989; US-431565.
 PA (IMMU-) IMMULOGIC PHARM COR.
 PI Geffer ML, Garman RD, Greenstein JL, Juo M, Rogers BL;
 PI Brauer AW;
 DR WPI; 91-164136/22.
 DR N-PSDB; Q11836.
 PT New pure covalently linked human T cell reactive feline protein -
 PT and modified peptide(s), used to reduce effects of cat allergens
 PT and to diagnose sensitivity to allergens.
 PS Claim 2; Fig 1; 70pp; English.
 CC Poly-A mRNA from cat parotid and mandibular glands was used to
 CC produce cDNA clones for both chain 1 and chain 2 of TRFP. These
 CC clones were then used to screen a cat genomic library. Chain 1
 CC exists in two forms having different leader sequences (A and B).
 CC The sequence can be used to express the protein and peptide derivs.
 CC which stimulate T-cells in persons allergic to cats. The peptides
 CC can be used to reduce/eliminate the allergic response partic. by
 CC modifcn. of lymphokine prodn. by the T-cells. They can also be
 CC used to identify epitopes responsible for sensitivity. The DNA can
 CC be used to detect comparable sequence in other species, and also
 CC for prodn. of modified forms of TRFP esp. showing reduced binding
 CC to IgE and thus reduced tendency to cause adverse reactions.
 CC See also R12120-R12123.
 SQ Sequence 94 AA;

DB 3; Score 195; Match 100.0%; Predicted No. 1.14e-13;
 Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 71 krdvdlfltgtptdeyveqvaqykalpv 97

Qy

1 KRVDLFLTGTPDEYVEQVAQYKALPV 27

RESULT 6

ID R36539 standard; Protein; 92 AA.
AC R36539;
DT 12-AUG-1993 (first entry)
DE TRFP chain 1 (with Leader A).
KW Human T cell reactive feline protein; TRFP; leader A; leader B;
KW epitope.
OS Felis.
FH Key Location/Qualifiers
FT Peptide 1..22
FT /label= leader_peptide
PN W09308280-A.
PD 29-APR-1993.
PF 16-OCT-1992; U08694.
PR 16-OCT-1991; US-777859.
PR 13-DEC-1991; US-807529.
PA (IMMU-) IMMULOGIC PHARM CORP.
PI Bond JF, Garman RD, Kuo M, Morgenstern JP, Morville M;
PI Rogers BL;
DR WPI; 93-152473/18.
DR N-PSDB; Q41556.
PT Recombitope peptide having T-cell stimulating activity - for the
PT diagnosis and treatment of sensitivity to protein allergens,
PT auto:antigens and protein antigens
PS Disclosure; Fig 1; 73pp; English.
CC Chains 1 and 2 of the TRFP have been recombinantly expressed in E.
CC coli and purified. T cell epitope studies using overlapping peptide
CC regions derived from the TRFP amino acids sequence were used to
CC identify multiple T cell epitopes in each chain of TRFP.
SQ Sequence 92 AA;

DB 7; Score 195; Match 100.0%; Predicted No. 1.14e-13;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 29 krddvdlfltgtpdeyveqvaqyk alpv 55

Qy 1 KRVDLFLTGTPDEYVEQVAQYKALPV 27

RESULT 7

ID R36540 standard; Protein; 88 AA.
AC R36540;
DT 12-AUG-1993 (first entry)
DE TRFP chain 1 (with Leader B).
KW Human T cell reactive feline protein; TRFP; leader A; leader B;
KW epitope.
OS Felis.
FH Key Location/Qualifiers
FT Peptide 1..18
FT /label= leader_peptide
PN W09308280-A.
PD 29-APR-1993.
PF 16-OCT-1992; U08694.
PR 16-OCT-1991; US-777859.
PR 13-DEC-1991; US-807529.
PA (IMMU-) IMMULOGIC PHARM CORP.
PI Bond JF, Garman RD, Kuo M, Morgenstern JP, Morville M;
PI Rogers BL;
DR WPI; 93-152473/18.
DR N-PSDB; Q41557.
PT Recombitope peptide having T-cell stimulating activity - for the
PT diagnosis and treatment of sensitivity to protein allergens,
PT auto:antigens and protein antigens
PS Disclosure; Fig 1; 73pp; English.
CC Chains 1 and 2 of the TRFP have been recombinantly expressed in E.
CC coli and purified. T cell epitope studies using overlapping peptide
CC regions derived from the TRFP amino acids sequence were used to
CC identify multiple T cell epitopes in each chain of TRFP.
SQ Sequence 88 AA;

PD 16-MAY-1991.
 PF 02-NOV-1990; U06548.
 PR 03-NOV-1989; US-431565.
 PA (IMMU-) IMMULOGIC PHARM COR.
 PI Gefter ML, Garman RD, Greenstein JL, Juo M, Rogers BL;
 PI Brauer AW;
 DR WPI; 91-164136/22.
 DR N-PSDB; Q11837.
 PT New pure covalently linked human T cell reactive feline protein -
 PT and modified peptide(s), used to reduce effects of cat allergens
 PT and to diagnose sensitivity to allergens.
 PS Claim 2; Fig 1; 70pp; English.
 CC Poly-A mRNA from cat parotid and mandibular glands was used to
 CC produce cDNA clones for both chain 1 and chain 2 of TRFP. These
 CC clones were then used to screen a cat genomic library. Chain 1
 CC exists in two forms having different leader sequences (A and B).
 CC The sequence can be used to express the protein and peptide derivs.
 CC which stimulate T-cells in persons allergic to cats. The peptides
 CC can be used to reduce/eliminate the allergic response partic. by
 CC modificn. of lymphokine prodn. by the T-cells. They can also be
 CC used to identify epitopes responsible for sensitivity. The DNA can
 CC be used to detect comparable sequence in other species, and also
 CC for prodn. of modified forms of TRFP esp. showing reduced binding
 CC to IgE and thus reduced tendency to cause adverse reactions.
 CC See also R12119-R12123.
 SQ Sequence 96 AA;

DB 3; Score 195; Match 100.0%; Predicted No. 1.14e-13;
 Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 33 krddvdlflgtgtpdeyveqvaqykalpv 59
 |||||
 Qy 1 KRDDVDFLTGTPDEYVEQVAQYKALPV 27

RESULT 10

ID R27368 standard; protein; 96 AA.
 AC R27368;
 DT 25-FEB-1993 (first entry)
 DE TRFP Chain #1 with C1 leader B sequence.
 KW T cell reactive feline protein; cat allergy; allergic; IgE;
 KW desensitizing;
 OS Felis domesticus.
 FH Key Location/Qualifiers
 FT Peptide 1..27
 FT /label= Leader B
 FT Protein 28..96
 FT /label= TRFP chain #1
 PN W09215613-A.
 PD 17-SEP-1992.
 PF 20-FEB-1992; U01344.
 PR 28-FEB-1991; US-662193.
 PA (IMMU-) IMMULOGIC PHARM CORP.
 PI Bond J, Kuo M;
 DR WPI; 92-331670/40.
 PT Modified human T-cell reactive feline protein - stimulates T-cell
 PT in individuals allergic to cats and shows reduced
 PT histamine-releasing properties
 PS Claim 1; Fig 1; 35pp; English.
 CC This sequence represents a modified human T-cell reactive feline
 CC protein which stimulates T-cells from an individual who is allergic
 CC to cats, but which interacts with human IgE to a lesser extent than
 CC does affinity purified TRFP. The protein is modified by treating
 CC with either a mild alkali (pH 12.5-13.5, KOH, NaOH, LiOH or tertiary
 CC amines) or an enzyme which removes O-linked groups (carbohydrate
 CC moieties). It is used in desensitizing people who are allergic to cats.

SQ Sequence 96 AA;

DB 5; Score 195; Match 100.0%; Predicted No. 1.14e-13;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 33 krdvdlfltgtptdegveqvaqykalpv 59
|||||
Qy 1 KRDVDLFLTGTPTDEYVEQVAQYKALPV 27

RESULT 11

ID R27367 standard; protein; 94 AA.
AC R27367;
DT 25-FEB-1993 (first entry)
DE TRFP Chain #1 with C1 leader A sequence.
KW T cell reactive feline protein.
OS Felis domesticus.
FH Key Location/Qualifiers
FT Peptide 1..25
FT /label= Leader A
FT Protein 25..94
FT /label= TRFP chain #1
PN W09215613-A.
PD 17-SEP-1992.
PF 20-FEB-1992; U01344.
PR 28-FEB-1991; US-662193.
PA (IMMU-) IMMULOGIC PHARM CORP.
PI Bond J, Kuo M;
DR WPI; 92-331670/40.
PT Modified human T-cell reactive feline protein - stimulates T-cell
PT in individuals allergic to cats and shows reduced
PT histamine-releasing properties
PS Claim 1; Fig 1; 35pp; English.
CC This sequence represents a modified human T-cell reactive feline
CC protein which stimulates T-cells from an individual who is allergic
CC to cats, but which interacts with human IgE to a lesser extent than
CC does affinity purified TRFP. The protein is modified by treating
CC with either a mild alkali (pH 12.5-13.5, KOH, NaOH, LiOH or tertiary
CC amines) or an enzyme which removes O-linked groups (carbohydrate
CC moieties). It is useful in desensitising people who are allergic to cats.
SQ Sequence 94 AA;

DB 5; Score 195; Match 100.0%; Predicted No. 1.14e-13;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 31 krdvdlfltgtptdegveqvaqykalpv 57
|||||
Qy 1 KRDVDLFLTGTPTDEYVEQVAQYKALPV 27

RESULT 12

ID R54202 standard; Protein; 422 AA.
AC R54202;
DT 18-NOV-1994 (first entry)
DE snaA gene product involved in streptogramin biosynthetic pathway.
KW Antibiotic; streptogramin; snaA; snaB; snaC; biosynthesis; enzyme;
KW biosynthetic pathway; Streptomyces pristinaespiralis.
OS Streptomyces pristinaespiralis.
PN FR2696189-A.
PD 01-APR-1994.
PF 25-SEP-1992; 011441.
PR 25-SEP-1992; FR-011441.
PA (RHON) RHONE POULENC RORER SA.
PI Blanc V, Blanche F, Crouzet J, Jacques N, Lacroix P;
PI Thibaut D, Zagorec M;
PI 11375_01-12001-1-1

DR N-PSDB; Q64202.
PT DNA involved in streptogramin antibiotic biosynthesis - for
PT prodn. or bio-conversion of streptogramin(s) or prodn. of
PT streptogramin intermediates, derivs. or hybrid antibiotics
PS Claim 21; Page 49-51; 83pp; French.
CC The *snaA* gene product is involved in the biosynthesis of
CC streptogramins, antibiotics active against Gram-positive bacteria.
CC The identification of the sequences encoding the enzymes involved
CC in the biosynthetic pathway means that they can be isolated and
CC manipulated. Mutant microorganisms in which a step in the
CC streptogramin biosynthetic pathway is blocked can be cultured to
CC produce streptogramin intermediates, which may later be converted
CC to streptogramin derivatives. Recombinant cells may also be used
CC for the bioconversion of streptogramins from one form to another or
CC for the production of hybrid antibiotics.
SQ Sequence 422 AA;

DB 10; Score 67; Match 33.3%; Predicted No. 8.62e+00;
Matches 6; Conservative 10; Mismatches 1; Indels 1; Gaps 1;

Db 370 nidfpylpgsaddfvdhv 387
::|: :| |::|::|::|
Qy 3 DVDL-FLTGTPDEYVEQV 19

RESULT 13

ID R47872 standard; Protein; 453 AA.
AC R47872;
DT 02-AUG-1994 (first entry)
DE Enzyme/biocatalyst which desulphurises a fossil fuel.
KW Enzyme; biocatalyst; fossil fuel; oxidation; cleavage;
KW organosulphur compounds; coal.
OS *Rhodococcus rhodochrous*.
PN W09401563-A.
PD 20-JAN-1994.
PF 09-JUL-1993; U06497.
PR 10-JUL-1992; US-911845.
PA (ENER-) ENERGY BIOSYSTEMS CORP.
PI Denome SA, Kovacevich BR, Piddington CS, Rambosek J;
PI Young KD;
DR WPI; 94-035068/04.
DR N-PSDB; Q55131.
PT DNA encoding a bio catalyst which desulphurises fossil fuels -
PT obtd. from *Rhodococcus rhodochrous* bacteria, used to produce
PT microorganisms which degrade organic sulphur cpds.
PS Disclosure; Page 72-73; 104pp; English.
CC Microorganisms transformed with the DNA encoding the
CC enzyme/biocatalyst can be used to produce the enzyme/biocatalyst for
CC the selective oxidative cleavage of carbon-sulphur bonds for
CC desulphurisation of fossil fuels which contain organosulphur
CC compounds.
SQ Sequence 453 AA;

DB 9; Score 65; Match 61.5%; Predicted No. 1.33e+01;
Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Db 406 flpgsydefvdqv 418
|| |::|::|::|
Qy 7 FLTGTPDEYVEQV 19

RESULT 14

ID R20746 standard; Protein; 234 AA.
AC R20746;
DT 28-MAY-1992 (first entry)
PS Patent R-20746 - As second approved ch. 101133

KW Receptor-type protein tyrosine phosphatase; cellular metabolism;
 KW cancer; diabetes.
 OS Homo sapiens.
 PN W09201050-A.
 PD 23-JAN-1992.
 PF 11-JUL-1991; U04892.
 PR 11-JUL-1990; US-551270.
 PR 26-FEB-1991; US-654188.
 PA (UYNY-) NEW YORK UNIV.
 PI Schlessinger J;
 DR WPI; 92-056865/07.
 PT Human receptor-type protein tyrosine phosphatase - has DNA
 PT encoding it and antibodies specific for it, useful for screening
 PT drugs affecting R-PTPase activity, and detection of mutant genes
 PS Claim 5; Fig 5B; 77pp; English.
 CC The amino acid sequence is that of human receptor-type protein
 CC tyrosine phosphatase (R-PTPase) beta second conserved phosphatase. It
 CC is useful in methods for screening drugs and other agents which are
 CC capable of activating or inhibiting the R-PTPase activity and thereby
 CC affecting major pathways of cellular metabolism. Activation of
 CC R-PTPases, leading to dephosphorylation would serve as a counter-
 CC regulatory mechanism to prevent or inhibit growth, and may serve as
 CC an endogenous regulatory mechanism against cancer. Mutation or
 CC dysregulation of this receptor/enzyme system may promote susceptibility
 CC to cancer, diabetes, or other diseases associated with alterations in
 CC cellular phosphotyrosine metabolism. It can be used to raise antibodies
 CC which can be used in immunoassays to determine the presence and amt.
 CC of R-PTPases, or in immunoelectron microscopy for in situ detection of
 CC R-PTPase. See also R20743-R20748.
 SQ Sequence 234 AA;

DB 4; Score 63; Match 38.9%; Predicted No. 2.03e+01;
 Matches 7; Conservative 7; Mismatches 4; Indels 0; Gaps 0;

Db 122 dfileatqddyvlevrhf 139
 |::| :| |::| :| :|
 Qy 5 DLFLTGTPEYVEQVAQY 22

RESULT 15

ID R13119 standard; Protein; 452 AA.
 AC R13119;
 DT 08-OCT-1991 (first entry)
 DE Phenylalanine hydroxylase.
 KW Hybrid; fusion; membrane translocation; binding region; HIV;
 KW infection; toxin; steroid; hormone; monoclonal antibody; antigen;
 KW diphtheria; exotoxin; phenylketonuria; cholera; interleukin; IL-2;
 KW protease; epidermal growth factor; ricin; tetanus; hexosaminidase;
 KW Shiga-like toxin A; SLT-A; PH; ligand; insulin; nuclease.
 OS Vibrio cholera.
 PN W09109871-A.
 PD 11-JUL-1991.
 PF 21-DEC-1990; U07619.
 PR 22-DEC-1989; US-456095.
 PR 14-JUN-1990; US-538276.
 PA (SERA-) SERAGEN INC.
 PI Murphy JR;
 DR WPI; 91-222845/30.
 DR N-PSDB; 012712.
 PT Hybrid molecules for targetting chemical entity to cell - have
 PT membrane trans-locating and cell binding-regions and used to
 PT treat HIV infection, genetic enzyme-deficiency disorders etc.
 PS Disclosure; Fig 13(1-3); 59pp; English.
 CC Hybrid molecules are produced by covalently linking
 CC (1) a portion (A) of the binding domain of a cell-binding ligand,
 CC allowing binding of the molecule to a target cell;

CC (2) a portion (B) of a translocation domain of a protein able to
 CC translocate (C) across the cell cytoplasmic membrane, and
 CC and (3) a portion (C) which is to be introduced into the cell.
 CC (A) is derived from a steroid or polypeptide hormone, a single-chain
 CC analogue of a monoclonal antibody able to bind an antigen expressed
 CC on the cell surface, or a polypeptide toxin.
 CC (B) is derived from a toxin (e.g. diphtheria toxin or Pseudomonas
 CC exotoxin A).
 CC (A) may be derived from insulin, interleukins 2, 3 or 6 or
 CC epidermal growth factor.
 CC Suitable enzymes in (C) include cholera toxin, ricin, tetanus toxin,
 CC hexosaminidase A, protease, nuclease, SLT-A, etc.
 CC Specified examples are CT-A/DT-B'/IL-2, SLTA/DT-B'/IL-2,
 CC ricin A/DT-B'/IL-2, HIVP-BP/DT-B'/IL-2 and the phenylalanine
 CC hydroxylase-DT-B' or their biologically active mutants.
 CC (CT-A= cholera toxin, DT-B'= truncated diphtheria toxin,
 CC SLTA= Shiga-like toxin A; HIVP-BP= HIV protease binding protein.
 CC See also 012710-12.
 SQ Sequence 452 AA;

DB 3; Score 63; Match 42.1%; Predicted No. 2.03e+01;
 Matches 8; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

Db 304 qeiglaslgapdeyiekla 322
 :: | :| | | | : |
 Qy 2 RDVDLFLTGPDEYVEQVA 20

Search completed: Fri Mar 24 07:42:29 1995
 Job time : 13 secs.

MASSIVE (TM)

Release 2.0 John F. Collins & S. S. Sturrock, Biocomputing Research Unit.
 Copyright (c) 1993, 1994 by University of Edinburgh, U.K.
 Distribution rights by IntelliGenetics, Inc.

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Mar 24 07:41:35 1995; MasPar time 5.07 Seconds
 119.562 Million cell updates/sec

Tabular output not generated.

Title: XUS-08-300-510-1
 Description: (1:27) from US08300510.pep
 Perfect Score: 195
 Sequence: 1 KRVDLFLTGPDEYVEQVAQYKALPV 27

Scoring table: PAM 150
 Gap 14

Searched: 75511 seqs, 22468834 residues

Database: pir43
 1 ANNO1
 2 ANNO2
 7 ANNO7

4 UNANNO1
5 UNANNO2
6 UNANNO3
7 UNANNO4
8 UNANNO5
9 UNANNO6
10 UNREV1
11 UNREV2
12 UNREV3

Statistics: Mean 30.206; Variance 58.847; scale 0.513

Predicted No. is the number of results expected by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	195	100.0	40	9	A53283	major cat allergen F	9.13e-22
2	189	96.9	88	9	JC1126	major allergen chain	1.27e-20
3	189	96.9	92	9	JC1136	major allergen chain	1.27e-20
4	74	37.9	1514	6	A29617	glutamate synthase (9.16e-01
5	71	36.4	703	5	A44983	heat shock protein 8	2.40e+00
6	70	35.9	85	11	S39326	auxin-induced mRNA -	3.29e+00
7	69	35.4	703	5	S08119	heat shock protein 8	4.50e+00
8	67	34.4	357	11	S33144	anthocyanidin hydrox	8.35e+00
9	67	34.4	452	5	A44888	heat shock protein 9	8.35e+00
10	66	33.8	1160	6	A45915	DNA-directed DNA pol	1.13e+01
11	66	33.8	1003	8	A38234	oxoglutarate dehydro	1.13e+01
12	66	33.8	336	10	S23484	xylZ protein - Pseud	1.13e+01
13	66	33.8	1160	1	DJEC3A	DNA-directed DNA pol	1.13e+01
14	66	33.8	336	6	C41659	benzoate 1,2-dioxyge	1.13e+01
15	65	33.3	1536	11	S39510	Glutamate Synthase (1.53e+01
16	65	33.3	1536	11	S31911	glutamate synthase -	1.53e+01
17	65	33.3	346	2	LUMS1	annexin I - mouse	1.53e+01
18	65	33.3	4969	9	A37113	ryanodine receptor,	1.53e+01
19	65	33.3	1955	2	AGCH	agrin precursor - ch	1.53e+01
20	64	32.8	275	4	A36690	sucrose alpha-glucos	2.07e+01
21	64	32.8	1556	5	D36793	hypothetical protein	2.07e+01
22	64	32.8	517	9	A49413	perilipin A - rat	2.07e+01
23	63	32.3	610	12	S12051	protein-tyrosine-pho	2.78e+01
24	63	32.3	332	3	Q0BE39	BGLF3 protein - huma	2.78e+01
25	63	32.3	1442	9	B48148	protein-tyrosine-pho	2.78e+01
26	63	32.3	246	1	WHRTF	phenylalanine 4-mono	2.78e+01
27	63	32.3	709	5	S03812	uvrB protein - Micro	2.78e+01
28	63	32.3	453	4	S15758	phenylalanine 4-mono	2.78e+01
29	63	32.3	476	9	B36065	protein-tyrosine-pho	2.78e+01
30	63	32.3	480	9	C36065	protein-tyrosine-pho	2.78e+01
31	63	32.3	453	4	A25321	phenylalanine 4-mono	2.78e+01
32	63	32.3	216	6	S25282	hypothetical protein	2.78e+01
33	63	32.3	1445	9	A48148	tyrosine phosphatase	2.78e+01
34	63	32.3	2307	9	A46700	receptor-type protei	2.78e+01
35	63	32.3	704	5	A26125	heat shock protein 9	2.78e+01
36	63	32.3	332	10	S33035	hypothetical protein	2.78e+01
37	63	32.3	452	1	WHHUF	phenylalanine 4-mono	2.78e+01
38	63	32.3	2314	9	A46151	protein-tyrosine-pho	2.78e+01
39	62	31.8	76	6	C47342	lct 3'-region hypoth	3.72e+01
40	61	31.3	764	8	A45321	protein-glutamine ga	4.97e+01
41	60	30.8	453	4	A42271	tryptophan 5-monooxy	6.61e+01
42	60	30.8	564	7	S30779	HCM1 protein - yeast	6.61e+01
43	60	30.8	2331	10	S32776	structural protein -	6.61e+01
44	60	30.8	411	5	A34526	ORF1 protein - Orgyi	6.61e+01
45	60	30.8	157	4	A00744	phenylalanine 4-mono	6.61e+01

ALIGNMENTS

RESULT 1

ENTRY A53283 #type fragment
 TITLE major cat allergen Fel d I alpha chain - cat (fragment)
 ORGANISM #formal_name Felis silvestris catus #common_name domestic cat
 DATE 12-May-1994 #sequence_revision 12-May-1994 #text_change
 12-May-1994
 ACCESSIONS A53283
 REFERENCE A53283
 #authors Duffort, D.A.; Carreira, J.; Nitti, G.; Polo, F.; Lombardero,
 M.
 #journal Mol. Immunol. (1991) 28:301-309
 #title Studies on the biochemical structure of the major cat
 allergen Felis domesticus I.
 #accession A53283
 ##status preliminary
 ##molecule_type protein
 ##residues 1-40 ##label DUF
 SUMMARY #length 40 #checksum 3032

DB 9; Score 195; Match 100.0%; Predicted No. 9.13e-22;
 Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 7 krdvdlfltgtptdeyveqvaqykalpv 33
 |||||
 Qy 1 KRVDLFLTGTPTDEYVEQVAQYKALPV 27

RESULT 2

ENTRY JC1126 #type complete
 TITLE major allergen chain 1 precursor B - cat
 ORGANISM #formal_name Felis silvestris catus #common_name domestic cat
 DATE 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change
 31-Dec-1993
 ACCESSIONS JC1126
 REFERENCE JC1126
 #authors Griffith, I.J.; Craig, S.; Pollock, J.; Yu, X.B.;
 Morgenstern, J.P.; Rogers, B.L.
 #journal Gene (1992) 113:263-268
 #title Expression and genomic structure of the genes encoding FdI,
 the major allergen from the domestic cat.
 #accession JC1126
 ##molecule_type DNA
 ##residues 1-88 ##label GRI

GENETICS

#gene Ch1
 #introns 17/1; 79/3

FEATURE

1-18 #domain signal sequence #status predicted #label SIG\
 19-88 #product major allergen chain 1 #status predicted #label
 MAT

SUMMARY #length 88 #molecular-weight 9586 #checksum 4095

DB 9; Score 189; Match 96.3%; Predicted No. 1.27e-20;
 Matches 26; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 25 krdvdlfltgtptdeyveqvaqynalp 51
 |||||
 Qy 1 KRVDLFLTGTPTDEYVEQVAQYKALPV 27

RESULT 3

ENTRY JC1126 #type complete

TITLE major allergen chain 1 precursor A - cat
 ORGANISM #formal_name Felis silvestris catus #common_name domestic cat
 DATE 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 31-Dec-1993
 ACCESSIONS JC1136
 REFERENCE JC1126
 #authors Griffith, I.J.; Craig, S.; Pollock, J.; Yu, X.B.; Morgenstern, J.P.; Rogers, B.L.
 #journal Gene (1992) 113:263-268
 #title Expression and genomic structure of the genes encoding FdI, the major allergen from the domestic cat.
 #accession JC1136
 ##molecule_type DNA
 ##residues 1-92 ##label GRI
 GENETICS
 #gene Ch1
 #introns 21/1; 83/3
 FEATURE
 1-22 #domain signal sequence #status predicted #label SIG\
 23-92 #product major allergen chain 1 #status predicted #label MAT
 SUMMARY #length 92 #molecular-weight 10072 #checksum 4988

DB 9; Score 189; Match 96.3%; Predicted No. 1.27e-20;
 Matches 26; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 29 krDVDLflTgTpdeYveQvAqYnAlpV 55
 |||||
 Qy 1 KRDVDLFLTGTPDEYVEQVAQYKALPV 27

RESULT 4
 ENTRY A29617 #type complete
 TITLE glutamate synthase (NADPH) (EC 1.4.1.13) large chain - Escherichia coli
 ORGANISM #formal_name Escherichia coli
 DATE 05-Jun-1988 #sequence_revision 05-Jun-1988 #text_change 31-Dec-1993
 ACCESSIONS A29617
 REFERENCE A91585
 #authors Oliver, G.; Gosset, G.; Sanchez-Pescador, R.; Lozoya, E.; Ku, L.M.; Flores, N.; Becerril, B.; Valle, F.; Bolivar, F.
 #journal Gene (1987) 60:1-11
 #title Determination of the nucleotide sequence for the glutamate synthase structural genes of Escherichia coli K-12.
 #cross-references MUID:88152492
 #contents K12
 #accession A29617
 ##molecule_type DNA
 ##residues 1-1514 ##label OLI
 ##note sequence not compared to nucleotide translation
 GENETICS
 #gene gltB
 #map_position 69
 KEYWORDS flavoprotein; glutamate biosynthesis; iron-sulfur protein; NADP; oxidoreductase
 SUMMARY #length 1514 #molecular-weight 166224 #checksum 756

DB 6; Score 74; Match 53.8%; Predicted No. 9.16e-01;
 Matches 7; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Db 1314 velyltgdandyv 1326
 |||||
 Qy 4 VDLFLTGTPDEYV 16

RESULT 5

ENTRY A44983 #type complete

TITLE heat shock protein 83 - Trypanosoma brucei

ORGANISM #formal_name Trypanosoma brucei

DATE 14-May-1993 #sequence_revision 14-May-1993 #text_change
30-Sep-1993

ACCESSIONS A44983

REFERENCE A44983

#authors Mottram, J.C.; Murphy, W.J.; Agabian, N.

#journal Mol. Biochem. Parasitol. (1989) 37:115-128

#title A transcriptional analysis of the Trypanosoma brucei hsp83
gene cluster.

#accession A44983

##status preliminary

##molecule_type DNA

##residues 1-703 ##label MOT

##cross-references GB:X14176

CLASSIFICATION #superfamily heat shock protein 90

SUMMARY #length 703 #molecular-weight 80729 #checksum 8300

DB 5; Score 71; Match 41.7%; Predicted No. 2.40e+00;

Matches 10; Conservative 7; Mismatches 5; Indels 2; Gaps 2;

Db 488 rrgmevlfntdpideyvmqqvkf 511

:|::||| ||||::||::

@y 1 KRDVD-LFLTGTPDEYV-EQVAQY 22

RESULT 6

ENTRY S39326 #type complete

TITLE auxin-induced mRNA - Arabidopsis thaliana

ORGANISM #formal_name Arabidopsis thaliana #common_name mouse-ear
cress

DATE 19-May-1994; #sequence_revision 19-May-1994; #text_change
19-May-1994

ACCESSIONS S39326

REFERENCE S39321

#authors Krivitzky, M.; Bonnet, R.; Jean-Jacques, I.; Kreis, M.;
Lecharny, A.

#submission submitted to the EMBL Data Library, December 1993

#accession S39326

##status preliminary

##residues 1-85 ##label KRI

##cross-references EMBL:Z29042

SUMMARY #length 85 #molecular-weight 9794 #checksum 7478

DB 11; Score 70; Match 40.9%; Predicted No. 3.29e+00;

Matches 9; Conservative 6; Mismatches 7; Indels 0; Gaps 0;

Db 22 emalklkgipyeyveeilnks 43

::||| ||||:::|:

@y 3 DVDLFLTGTPDEYVEQVAQYKA 24

RESULT 7

ENTRY S08119 #type complete

TITLE heat shock protein 83 - Trypanosoma brucei brucei

ORGANISM #formal_name Trypanosoma brucei brucei

DATE 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change
18-Jun-1993

ACCESSIONS S08119

REFERENCE S08119

#authors Mottram, J.; Murphy, W.; Agabian, N.

#submission submitted to the EMBL Data Library, January 1989

#accession S08119

##molecule_type DNA

##residues 1-703 ##label MOT
##cross-references EMBL:X14176

GENETICS

#gene hsp83
CLASSIFICATION #superfamily heat shock protein 90
SUMMARY #length 703 #molecular-weight 80715 #checksum 8246

DB 5: Score 69; Match 41.7%; Predicted No. 4.50e+00;
Matches 10; Conservative 7; Mismatches 5; Indels 2; Gaps 2;

Db 488 rrgnevlfmtdpideyvmqqvkdf 511
:|::||| |||:|:|:
Qy 1 KRDVD-LFLTGTPEYV-EQVAQY 22

RESULT 8

ENTRY S33144 #type complete
TITLE anthocyanidin hydroxylase - apple tree
ORGANISM #formal_name Malus sp. #common_name apple tree
DATE 22-Nov-1993; #sequence_revision 22-Nov-1993; #text_change 22-Nov-1993

ACCESSIONS S33144
REFERENCE S33144
#authors Davies, K.M.
#submission submitted to the EMBL Data Library, March 1993
#accession S33144
##status preliminary
##residues 1-357 ##label DAV
##cross-references EMBL:X71360
SUMMARY #length 357 #molecular-weight 40332 #checksum 9659

DB 11: Score 67; Match 40.9%; Predicted No. 8.35e+00;
Matches 9; Conservative 5; Mismatches 7; Indels 1; Gaps 1;

Db 156 krdlsiw-pqtpadyieataey 176
|||: : ||:|:| |:
Qy 1 KRVDLFLTGTPEYVEQVAQY 22

RESULT 9

ENTRY A44888 #type fragment
TITLE heat shock protein 90 - Leishmania donovani (fragment)
ALTERNATE_NAMES heat shock protein, 90K
ORGANISM #formal_name Leishmania donovani
DATE 17-Feb-1994 #sequence_revision 17-Feb-1994 #text_change 17-Feb-1994
ACCESSIONS A44888
REFERENCE A44888
#authors de Andrade, C.R.; Kirchhoff, L.V.; Donelson, J.E.; Otsu, K.
#journal J. Clin. Microbiol. (1992) 30:330-335
#title Recombinant Leishmania Hsp90 and Hsp70 are recognized by sera from visceral leishmaniasis patients but not Chagas' disease patients.
#cross-references MUID:92165942
#contents strain Sudan S1
#accession A44888
##status preliminary
##molecule_type mRNA
##residues 1-452 ##label DE1
##cross-references NCBI:83989; NCBI:83991
##note sequence extracted from NCBI backbone
CLASSIFICATION #superfamily heat shock protein 90
KEYWORDS heat shock protein; phosphoprotein
SUMMARY #length 452 #checksum 8278

DB 5: Score 67; Match 41.7%; Predicted No. 8.35e+00;

Matches 10; Conservative 7; Mismatches 5; Indels 2; Gaps 2;

Db 237 rrglevlfntepideyvmqqvkdf 260
:|::||| | ||| :|| :|
Qy 1 KRDVD-LFLTGTPEYV-EQVAQY 22

RESULT 10

ENTRY A45915 #type complete
TITLE DNA-directed DNA polymerase (EC 2.7.7.7) III alpha chain -
Salmonella typhimurium
ORGANISM #formal_name Salmonella typhimurium
DATE 10-Mar-1994 #sequence_revision 10-Mar-1994 #text_change
18-Nov-1994
ACCESSIONS A45915
REFERENCE A45915
#authors Lancy, E.D.; Lifshits, M.R.; Munson, P.; Maurer, R.
#journal J. Bacteriol. (1989) 171:5581-5586
#title Nucleotide sequence of dnaE, the gene for the polymerase
subunit of DNA polymerase III in Salmonella typhimurium,
and a variant that facilitates growth in the absence of
another polymerase subunit.
#accession A45915
##status preliminary
##molecule_type DNA
##residues 1-1160 ##label LAN
##cross-references GB:M29701
##note translation of nucleotide sequence not given
KEYWORDS glycosidase; hydrolase; nucleotidyltransferase
SUMMARY #length 1160 #molecular-weight 130118 #checksum 9028

DB 6; Score 66; Match 35.0%; Predicted No. 1.13e+01;
Matches 7; Conservative 8; Mismatches 4; Indels 1; Gaps 1;

Db 957 lglyltghpinqylkeiery 976
:|:|||| |::|: ::|
Qy 4 VDLFLTGTPEYVEQVAQY 22

RESULT 11

ENTRY A38234 #type complete
TITLE oxoglutarate dehydrogenase (lipoamide) (EC 1.2.4.2) precursor
- human
ALTERNATE_NAMES 2-oxoglutarate:lipoamide 2-oxoreductase;
alpha-ketoglutarate dehydrogenase
ORGANISM #formal_name Homo sapiens #common_name man
DATE 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change
31-Dec-1993
ACCESSIONS A38234
REFERENCE A38234
#authors Koike, K.; Urata, Y.; Goto, S.
#journal Proc. Natl. Acad. Sci. U.S.A. (1992) 89:1963-1967
#title Cloning and nucleotide sequence of the cDNA encoding human
2-oxoglutarate dehydrogenase (lipoamide).
#cross-references MUID:92179301
#contents fetal liver
#accession A38234
##molecule_type mRNA
##residues 1-1003 ##label KOI
##cross-references GB:D10523; NCBIP:87352
##note sequence extracted from NCBI backbone
KEYWORDS mitochondrion; oxidoreductase; thiamine pyrophosphate;
tricarboxylic acid cycle

FEATURE 1-40 #domain transit peptide (mitochondrion) #status
predicted #label TMD1

41-1003 #product oxoglutarate dehydrogenase (lipoamide) #status
predicted #label MAT
SUMMARY #length 1003 #molecular-weight 113239 #checksum 617

DB 8: Score 66; Match 46.7%; Predicted No. 1.13e+01;
Matches 7; Conservative 6; Mismatches 2; Indels 0; Gaps 0;

Db 47 epflsgtssnyveem 61
: ||: ||: : ||: ||:
Qy 5 DLFLTGTPDEYVEQV 19

RESULT 12

ENTRY S23484 #type complete
TITLE xylZ protein - Pseudomonas putida plasmid pWWO
ORGANISM #formal_name Pseudomonas putida
DATE 22-Nov-1993; #sequence_revision 22-Nov-1993; #text_change
22-Nov-1993
ACCESSIONS S23484
REFERENCE S23477
#authors Neidle, E.L.; Hartnett, C.; Ornstom, L.N.; Bairoch, A.;
Rekik, M.; Harayama, S.
#journal Eur. J. Biochem. (1992) 204:113-120
#title Cis-diol dehydrogenases encoded by the TOL pWWO plasmid xylL
gene and the Acinetobacter calcoaceticus chromosomal bend
gene are members of the short-chain alcohol dehydrogenase
superfamily.
#cross-references MUID:92155191
#accession S23484
##status preliminary
##residues 1-336 ##label NEI
##cross-references EMBL:M64747
SUMMARY #length 336 #molecular-weight 36220 #checksum 5068

DB 10: Score 66; Match 50.0%; Predicted No. 1.13e+01;
Matches 10; Conservative 4; Mismatches 5; Indels 1; Gaps 1;

Db 298 evdiylcgpppm-veavsqy 316
: ||: ||: ||: ||: ||: ||:
Qy 3 DVDLFLTGTPDEYVEQVAQY 22

RESULT 13

ENTRY DJEC3A #type complete
TITLE DNA-directed DNA polymerase (EC 2.7.7.7) III alpha chain -
Escherichia coli
ORGANISM #formal_name Escherichia coli
DATE 31-Dec-1988 #sequence_revision 31-Dec-1988 #text_change
12-May-1994
ACCESSIONS C28390; A37441
REFERENCE A91855
#authors Tomaszewicz, H.G.; McHenry, C.S.
#journal J. Bacteriol. (1987) 169:5735-5744
#title Sequence analysis of the Escherichia coli dnaE gene.
#cross-references MUID:88058791
#accession C28390
##molecule_type DNA
##residues 1-1160 ##label TOM
##note the nucleotide sequence has been corrected in reference
A37441
REFERENCE A37441
#authors Tomaszewicz, H.G.; McHenry, C.S.
#journal J. Bacteriol. (1991) 173:4549
#contents erratum
#accession A37441
##molecule_type DNA

##residues 156-183 ##label T02
##cross-references GB:M19334
COMMENT This protein is the catalytic component of the DNA polymerase III core (pol III, composed of alpha, epsilon, and theta chains) that can repair short gaps created by nuclease in duplex DNA. For efficient replication of the long, single-stranded templates, pol III requires the auxiliary chains beta, gamma, and delta.

GENETICS

#gene dnaE
#map_position 4 min
CLASSIFICATION #superfamily DNA-directed DNA polymerase III alpha chain
KEYWORDS DNA replication; nucleotidyltransferase
SUMMARY #length 1160 #molecular-weight 129903 #checksum 8714

DB 1; Score 66; Match 35.0%; Predicted No. 1.13e+01;
Matches 7; Conservative 8; Mismatches 4; Indels 1; Gaps 1;

Db 957 lglyltghpinqylkeiery 976
: |||| | :|: :| :|
Qy 4 VDLFLTGTP-DEYVEQVAQY 22

RESULT 14

ENTRY C41659 #type complete
TITLE benzoate 1,2-dioxygenase (EC 1.14.12.10) XylZ protein -
Pseudomonas putida plasmid pWWO
ORGANISM #formal_name Pseudomonas putida
DATE 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change
30-Sep-1993
ACCESSIONS C41659
REFERENCE A41659
#authors Harayana, S.; Rekik, M.; Bairoch, A.; Neidle, E.L.; Ornston,
L.N.
#journal J. Bacteriol. (1991) 173:7540-7548
#title Potential DNA slippage structures acquired during
evolutionary divergence of Acinetobacter calcoaceticus
chromosomal benABC and Pseudomonas putida TOL pWWO plasmid
xylXYZ, genes encoding benzoate dioxygenases.
#cross-references MUID:92041666
#accession C41659
##status preliminary
##molecule_type DNA
##residues 1-336 ##label HAR
##cross-references GB:M64747

GENETICS

#genome plasmid
SUMMARY #length 336 #molecular-weight 36220 #checksum 5068

DB 6; Score 66; Match 50.0%; Predicted No. 1.13e+01;
Matches 10; Conservative 4; Mismatches 5; Indels 1; Gaps 1;

Db 298 evdiylcgpppn-veavsqy 316
: ||:| | | || |:
Qy 3 DVDLFLTGTPDEYVEQVAQY 22

RESULT 15

ENTRY S39510 #type complete
TITLE Glutamate Synthase (EC 1.4.7.1) - Antithamnion sp.
ORGANISM #formal_name Antithamnion sp.
DATE 19-May-1994; #sequence_revision 19-May-1994; #text_change
19-May-1994
ACCESSIONS S39510
REFERENCE S39510
#authors Valentin, K.; Kostrzewa, M.; Zetsche, K.
#cross-references GB:M64747

#title Glutamate synthase is plastid-encoded in a red alga:
implications for the evolution of glutamate synthases.
#accession S39510
##status preliminary
##residues 1-1536 ##label VAL
##cross-references EMBL:Z21705
SUMMARY #length 1536 #molecular-weight 171111 #checksum 6689

DB 11; Score 65; Match 33.3%; Predicted No. 1.53e+01;
Matches 5; Conservative 6; Mismatches 4; Indels 0; Gaps 0;

Db 1340 kgiHlylkgeandyv 1354
: : | : | : : : |
Qy 2 RDVDLFLTGTPEYV 16

Search completed: Fri Mar 24 07:41:57 1995
Job time : 22 secs.

MASPAR (TM)

Release 2.0 John F. Collins & S. S. Sturrock, Biocomputing Research Unit.
Copyright (c) 1993, 1994 by University of Edinburgh, U.K.
Distribution rights by IntelliGenetics, Inc.

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Mar 24 07:41:03 1995; MasPar time 3.49 Seconds
109.451 Million cell updates/sec

Tabular output not generated.

Title: >US-08-300-510-1
Description: (1:27) from US08300510.pep
Perfect Score: 195
Sequence: 1 KRVDLFLTGTPEYVEQVAQYKALPV 27

Scoring table: PAM 150
Gap 14

Searched: 40292 seqs, 14147368 residues

Database: swiss-prot30
1 SPT1
2 SPT2
3 SPT3
4 SPT4
5 SPT5
6 SPT6
7 SPT7

Statistics: Mean 31.852; Variance 49.755; scale 0.640

Predicted No. is the number of results expected by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

Result No.	Score	% Match	Query Length	DB ID	Description	Pred. No.
1	195	100.0	88	2 FELB_FELCA	MAJOR ALLERGEN I POLY	3.32e-27
2	195	100.0	92	2 FELA_FELCA	MAJOR ALLERGEN I POLY	3.32e-27
3	74	37.9	1514	3 GLTB_ECOLI	GLUTAMATE SYNTHASE (N	1.41e-01
4	69	35.4	703	3 HS83_TRYBB	HEAT SHOCK PROTEIN 83	9.34e-01
5	67	34.4	452	3 HS90_LEIDO	HEAT SHOCK PROTEIN 90	1.94e+00
6	66	33.8	336	7 XYLZ_PSEPU	TOLUATE 1,2-DIOXYGENA	2.78e+00
7	66	33.8	1160	2 DP3A_ECOLI	DNA POLYMERASE III, A	2.78e+00
8	66	33.8	1003	5 ODO1_HUMAN	2-OXOGLUTARATE DEHYDR	2.78e+00
9	66	33.8	1160	2 DP3A_SALTY	DNA POLYMERASE III, A	2.78e+00
10	65	33.3	1955	1 AGRI_CHICK	AGRIN PRECURSOR.	3.97e+00
11	65	33.3	345	1 ANX1_MOUSE	ANNEXIN I (LIPOCORTIN	3.97e+00
12	65	33.3	4969	6 RYNC_RABIT	RYANODINE RECEPTOR, C	3.97e+00
13	65	33.3	1536	3 GLSF_ANTSP	FERREDOXIN-DEPENDENT	3.97e+00
14	64	32.8	1556	7 VG67_HSVI1	HYPOTHETICAL GENE 67	5.64e+00
15	64	32.8	917	6 SUIS_RAT	SUCRASE-ISOMALTASE, I	5.64e+00
16	63	32.3	332	7 UL95_EBV	HYPOTHETICAL PROTEIN	7.98e+00
17	63	32.3	1442	5 PTPG_MOUSE	PROTEIN-TYROSINE PHOS	7.98e+00
18	63	32.3	704	3 HS85_TRYCR	HEAT SHOCK LIKE 85 KD	7.98e+00
19	63	32.3	452	5 PH4H_HUMAN	PHENYLALANINE-4-HYDRO	7.98e+00
20	63	32.3	709	7 UVRB_MICLU	EXCINUCLEASE ABC SUBU	7.98e+00
21	63	32.3	216	7 YHCA_ECOLI	HYPOTHETICAL 24.6 KD	7.98e+00
22	63	32.3	2314	5 PTPZ_HUMAN	PROTEIN-TYROSINE PHOS	7.98e+00
23	63	32.3	453	5 PH4H_MOUSE	PHENYLALANINE-4-HYDRO	7.98e+00
24	63	32.3	1445	5 PTPG_HUMAN	PROTEIN-TYROSINE PHOS	7.98e+00
25	63	32.3	453	5 PH4H_RAT	PHENYLALANINE-4-HYDRO	7.98e+00
26	62	31.8	192	7 YC08_YEAST	HYPOTHETICAL 21.1 KD	1.12e+01
27	62	31.8	922	4 LCN2_LACLA	LACTICIN 481/LACTOCOC	1.12e+01
28	61	31.3	721	3 HS90_THEPA	HEAT SHOCK PROTEIN 90	1.58e+01
29	60	30.8	1147	4 MYSB_ACACA	MYOSIN HEAVY CHAIN IB	2.20e+01
30	60	30.8	453	5 PH4H_DROME	PHENYLALANINE-4-HYDRO	2.20e+01
31	60	30.8	411	7 VP48_NPVOP	P48 PROTEIN.	2.20e+01
32	60	30.8	2331	6 RRPL_MABVP	RNA-DIRECTED RNA POLY	2.20e+01
33	60	30.8	564	3 HCM1_YEAST	HCM1 PROTEIN.	2.20e+01
34	60	30.8	443	2 DACA_BACSU	D-ALANYL-D-ALANINE CA	2.20e+01
35	59	30.3	1509	4 MYSN_ACACA	MYOSIN II HEAVY CHAIN	3.06e+01
36	59	30.3	701	3 HS83_LEIAM	HEAT SHOCK PROTEIN 83	3.06e+01
37	59	30.3	389	1 ALR_BACSU	ALANINE RACEMASE (EC	3.06e+01
38	59	30.3	1202	6 RPM2_YEAST	MITOCHONDRIAL RIBONUC	3.06e+01
39	58	29.7	249	4 MYPO_CHICK	MYELIN P0 PROTEIN PRE	4.23e+01
40	58	29.7	689	7 YR40_BACSU	HYPOTHETICAL 78.8 KD	4.23e+01
41	58	29.7	308	7 YPX1_BLVJ	HYPOTHETICAL PXBL-I P	4.23e+01
42	58	29.7	453	6 TBB2_GEOCN	TUBULIN BETA-2 CHAIN.	4.23e+01
43	58	29.7	695	2 EFG_SYNY3	ELONGATION FACTOR G (4.23e+01
44	58	29.7	522	1 CEA1_ECOLI	COLICIN E1 PROTEIN.	4.23e+01
45	58	29.7	869	1 CFAC_ECOLI	CFA/I FIMBRIAL SUBUNI	4.23e+01

ALIGNMENTS

```

RESULT 1
ID FELB_FELCA STANDARD; PRT; 88 AA.
AC P30439;
DT 01-APR-1993 (REL. 25, CREATED)
DT 01-APR-1993 (REL. 25, LAST SEQUENCE UPDATE)
DT 01-JUN-1994 (REL. 29, LAST ANNOTATION UPDATE)
DE MAJOR ALLERGEN I POLYPEPTIDE CHAIN 1 MINOR FORM PRECURSOR (FEL D I)
DE (CAT-1) (AG 4).
GN CHI.
OS FELIS CATUS (CAT).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; CARNIVORA.

```


RP SEQUENCE FROM N.A., AND SEQUENCE OF 19-88.
 RM 92052157
 RA MORGENSTERN J.P., GRIFFITH I.J., BRAUER A.W., ROGERS B.L.,
 RA BOND J.F., CHAPMAN M.D., KUO M.-C.;
 RL PROC. NATL. ACAD. SCI. U.S.A. 88:9690-9694(1991).
 RN [2]
 RP SEQUENCE FROM N.A.
 RM 92241678
 RA GRIFFITH I.J., CRAIG S., POLLOCK J., YU X.-B., MORGENSTERN J.P.,
 RA ROGERS B.L.,
 RL GENE 113:263-268(1992).
 RN [3]
 RP SEQUENCE OF 19-58, AND CHARACTERIZATION.
 RM 91287714
 RA DUFFORT D.A., CARREIRA J., NITTI G., POLO F., LOMBARDEO M.;
 RL MOL. IMMUNOL. 28:301-309(1991).
 RN [4]
 RP CHARACTERIZATION.
 RA LEITERMANN K., OHMAN J.L. JR.;
 RL J. ALLERGY CLIN. IMMUNOL. 74:147-153(1991).
 CC -!- DISEASE: MAJOR ALLERGEN PRODUCED BY THE DOMESTIC CAT.
 CC -!- SUBUNIT: HETEROTETRAMER COMPOSED OF TWO NON-COVALENTLY LINKED
 CC DISULFIDE-LINKED HETERODIMER OF CHAINS 1 AND 2.
 CC -!- TISSUE SPECIFICITY: SALIVA, AND SEBACEOUS GLANDS.
 CC -!- ALTERNATIVE PRODUCTS: USAGE OF TWO DIFFERENT INITIATOR MET ARE
 CC RESPONSIBLE FOR THE PRODUCTION OF TWO FORMS OF THE SIGNAL SEQUENCE
 CC OF THIS ALLERGEN SUBUNIT.
 CC -!- SIMILARITY: TO UTEROGLOBIN.
 DR EMBL; M74953; FDFELDIB.
 DR PIR; JC1126; JC1126.
 DR PROSITE; PS00403; UTEROGLOBIN_1.
 DR PROSITE; PS00404; UTEROGLOBIN_2.
 KW ALLERGEN; SIGNAL; ALTERNATIVE SPLICING.
 FT SIGNAL 1 18
 FT CHAIN 19 88 MAJOR ALLERGEN I POLYPEPTIDE CHAIN 1.
 FT DISULFID 21 21 INTERCHAIN (POTENTIAL).
 FT DISULFID 88 88 INTERCHAIN (POTENTIAL).
 FT VARIANT 47 47 K -> N.
 FT CONFLICT 78 78 L -> V (IN REF. 2).
 SQ SEQUENCE 88 AA; 9614 MW; 39445 CN;

DB 2; Score 195; Match 100.0%; Predicted No. 3.32e-27;
 Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 25 krdvdlfltgtptdegveqvaqyk alpv 51
 |||||
 Qy 1 KRDVDLFLTGTPTDEYVEQVAQYKALPV 27

RESULT 2
 ID FELA_FELCA STANDARD; PRT; 92 AA.
 AC P30438;
 DT 01-APR-1993 (REL. 25, CREATED)
 DT 01-APR-1993 (REL. 25, LAST SEQUENCE UPDATE)
 DT 01-JUN-1994 (REL. 29, LAST ANNOTATION UPDATE)
 DE MAJOR ALLERGEN I POLYPEPTIDE CHAIN 1 MAJOR FORM PRECURSOR (FEL D I)
 DE (CAT-1) (AG 4).
 GN CH1.
 OS FELIS CATUS (CAT).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
 OC EUTHERIA; CARNIVORA.
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 23-92.
 RC TISSUE=SALIVARY GLAND;
 RM 92052157
 RA MORGENSTERN J.P., GRIFFITH I.J., BRAUER A.W., ROGERS B.L.,

RA BOND J.F., CHAPMAN M.D., KUD M.-C.;
 RL PROC. NATL. ACAD. SCI. U.S.A. 88:9690-9694(1991).
 RN [2]
 RP SEQUENCE FROM N.A.
 RM 92241678
 RA GRIFFITH I.J., CRAIG S., POLLOCK J., YU X.-B., MORGENSTERN J.P.,
 RA ROGERS B.L.,
 RL GENE 113:263-268(1992).
 RN [3]
 RP SEQUENCE OF 23-62, AND CHARACTERIZATION.
 RM 91287714
 RA DUFFORT D.A., CARREIRA J., NITTI G., POLO F., LOMBARDEO M.;
 RL MOL. IMMUNOL. 28:301-309(1991).
 RN [4]
 RP CHARACTERIZATION.
 RA LEITERMANN K., OHMAN J.L. JR.;
 RL J. ALLERGY CLIN. IMMUNOL. 74:147-153(1991).
 CC -!- DISEASE: MAJOR ALLERGEN PRODUCED BY THE DOMESTIC CAT.
 CC -!- SUBUNIT: HETEROTETRAMER COMPOSED OF TWO NON-COVALENTLY LINKED
 CC DISULFIDE-LINKED HETERODIMER OF CHAINS 1 AND 2.
 CC -!- TISSUE SPECIFICITY: SALIVA, AND SEBACEOUS GLANDS.
 CC -!- ALTERNATIVE PRODUCTS: USAGE OF TWO DIFFERENT INITIATOR MET ARE
 CC RESPONSIBLE FOR THE PRODUCTION OF TWO FORMS OF THE SIGNAL SEQUENCE
 CC OF THIS ALLERGEN SUBUNIT.
 CC -!- SIMILARITY: TO UTEROGLOBIN.
 DR EMBL; M74952; FDFELDI.
 DR PIR; JC1136; JC1136.
 DR PROSITE; PS00403; UTEROGLOBIN_1.
 DR PROSITE; PS00404; UTEROGLOBIN_2.
 KW ALLERGEN; SIGNAL; ALTERNATIVE SPLICING.
 FT SIGNAL 1 22
 FT CHAIN 23 92 MAJOR ALLERGEN I POLYPEPTIDE CHAIN 1.
 FT DISULFID 25 25 INTERCHAIN (POTENTIAL).
 FT DISULFID 92 92 INTERCHAIN (POTENTIAL).
 FT VARIANT 51 51 K -> N.
 FT CONFLICT 5 5 R -> C (IN REF. 2).
 FT CONFLICT 18 18 W -> S (IN REF. 2).
 FT CONFLICT 82 82 L -> V (IN REF. 2).
 SQ SEQUENCE 92 AA; 10252 MW; 43206 CN;

DB 2; Score 195; Match 100.0%; Predicted No. 3.32e-27;
 Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 29 krdvdlfltgtpdeyveqvaykalpv 55
 |||||
 Qy 1 KRDVDLFLTGTGPDEYVEQVAQYKALPV 27

RESULT 3
 ID GLTB_ECOLI STANDARD; PRT: 1514 AA.
 AC P09831;
 DT 01-MAR-1989 (REL. 10, CREATED)
 DT 01-MAR-1989 (REL. 10, LAST SEQUENCE UPDATE)
 DT 01-OCT-1994 (REL. 30, LAST ANNOTATION UPDATE)
 DE GLUTAMATE SYNTHASE (NADPH) LARGE CHAIN PRECURSOR (EC 1.4.1.13)
 DE (NADPH-GOGAT).
 GN GLTB.
 OS ESCHERICHIA COLI.
 OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; FACULTATIVELY ANAEROBIC RODS;
 OC ENTEROBACTERIACEAE.
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC STRAIN=K12;
 RM 88152492
 RA OLIVER G., GOSSET G., SANCHEZ-PESCADOR R., LOZOYA E., KU L.M.,
 RA

RL GENE 60:1-11(1987).
 RN [2]
 RP DISCUSSION OF SEQUENCE.
 RM 89098858
 RA GOSSET G., MERINO E., RECILLAS F., OLIVER G., BECERRIL B., BOLIVAR F.;
 RL PROTEIN SEQ. DATA ANAL. 2:9-16(1989).
 CC -!- CATALYTIC ACTIVITY: 2 L-GLUTAMATE + NADP(+) = L-GLUTAMINE +
 CC 2-OXOGLUTARATE + NADPH.
 CC -!- PATHWAY: NITROGEN METABOLISM, GLUTAMATE BIOSYNTHESIS.
 CC THE CATALYZED REACTION BRINGS TOGETHER THE NITROGEN AND
 CC CARBON METABOLISM.
 CC -!- COFACTOR: IRON-SULFUR; FAD AND FMN FLAVOPROTEIN.
 CC -!- SUBUNIT: AGGREGATE OF 4 CATALYTICAL ACTIVE HETERODIMERS,
 CC CONSISTING OF A LARGE AND A SMALL SUBUNIT.
 CC -!- GLUTAMINE BINDS TO THE LARGE SUBUNIT AND TRANSFERS THE AMIDO GROUP
 CC TO 2-OXO-GLUTAMATE THAT APPARENTLY BINDS TO THE SMALL SUBUNIT.
 CC -!- SIMILARITY: TO OTHER GLUTAMATE SYNTHASES.
 DR EMBL; M18747; ECGLTB.
 DR PIR; A29617; A29617.
 DR ECOGENE; EG10403; GLTB.
 KW OXIDOREDUCTASE; IRON-SULFUR; FLAVOPROTEIN; FAD; FMN;
 KW GLUTAMATE BIOSYNTHESIS; SIGNAL.
 FT SIGNAL 1 42
 FT CHAIN 43 1514 GLUTAMATE SYNTHASE LARGE CHAIN.
 FT NP_BIND 1077 1134 FMN (BY SIMILARITY).
 SQ SEQUENCE 1514 AA; 166225 MW; 11266241 CN;

DB 3; Score 74; Match 53.8%; Predicted No. 1.41e-01;
 Matches 7; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Db 1314 velyltgdandyv 1326
 |:|:| | : : | |
 Qy 4 VDLFLTGTPDEYV 16

RESULT 4
 ID HS83_TRYBB STANDARD; PRT; 703 AA.
 AC P12861;
 DT 01-OCT-1989 (REL. 12, CREATED)
 DT 01-OCT-1989 (REL. 12, LAST SEQUENCE UPDATE)
 DT 01-OCT-1993 (REL. 27, LAST ANNOTATION UPDATE)
 DE HEAT SHOCK PROTEIN 83.
 GN HSP83.
 OS TRYPANOSOMA BRUCEI BRUCEI.
 OC EUKARYOTA; PROTOZOA; SARCOMASTIGOPHORA; MASTIGOPHORA; KINETOPLASTIDA;
 OC TRYPANOSOMATIDAE.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=LSTAR SERODEME;
 RM 90136708
 RA MOTTRAM J., MURPHY W., AGABIAN N.;
 RL MOL. BIOCHEM. PARASITOL. 37:115-128(1989).
 CC -!- FUNCTION: MOLECULAR CHAPERONE. HAS ATPASE ACTIVITY.
 CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
 CC -!- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN HSP90 FAMILY.
 DR EMBL; X14176; TBHSP83.
 DR PIR; S08119; S08119.
 DR PROSITE; PS00298; HSP90.
 KW CHAPERONE; ATP-BINDING; HEAT SHOCK.
 SQ SEQUENCE 703 AA; 80715 MW; 2466880 CN;

DB 3; Score 69; Match 41.7%; Predicted No. 9.34e-01;
 Matches 10; Conservative 7; Mismatches 5; Indels 2; Gaps 2;

Db 488 rrgmevlfntdpideyvmqqvkdf 511

Qy

1 KRDVD-LFLTGTPEYV-EQVAQY 22

RESULT 5

ID HS90_LEIDO STANDARD; PRT; 452 AA.
 AC P27890;
 DT 01-AUG-1992 (REL. 23, CREATED)
 DT 01-AUG-1992 (REL. 23, LAST SEQUENCE UPDATE)
 DT 01-OCT-1993 (REL. 27, LAST ANNOTATION UPDATE)
 DE HEAT SHOCK PROTEIN 90 (HSP 90) (FRAGMENT).
 GN HSP90.
 OS LEISHMANIA DONOVANI.
 OC EUKARYOTA; PROTOZOA; SARCOMASTIGOPHORA; MASTIGOPHORA; KINETOPLASTIDA;
 OC TRYPANOSOMATIDAE.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SUDAN S1;
 RM 92165942
 RA DE ANDRADE C.R., KIRCHHOFF L.V., DONELSON J.E., OTSU K.;
 RL J. CLIN. MICROBIOL. 30:330-335(1992).
 CC -!- FUNCTION: MOLECULAR CHAPERONE. HAS ATPASE ACTIVITY
 CC (BY SIMILARITY).
 CC -!- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN HSP90 FAMILY.
 DR EMBL; M73492; LDHSP90.
 DR PIR; A44888; A44888.
 DR PROSITE; PS00298; HSP90.
 KW CHAPERONE; ATP-BINDING; HEAT SHOCK.
 FT NON_TER 1 1
 SQ SEQUENCE 452 AA; 52691 MW; 1061521 CN;

DB 3; Score 67; Match 41.7%; Predicted No. 1.94e+00;
 Matches 10; Conservative 7; Mismatches 5; Indels 2; Gaps 2;

Db 237 rrglevlfmtepidayvmqqvkdf 260
 :| :: ||:| |||| :|| ::
 Qy 1 KRDVD-LFLTGTPEYV-EQVAQY 22

RESULT 6

ID XYLZ_PSEPU STANDARD; PRT; 336 AA.
 AC P23101;
 DT 01-NOV-1991 (REL. 20, CREATED)
 DT 01-NOV-1991 (REL. 20, LAST SEQUENCE UPDATE)
 DT 01-AUG-1992 (REL. 23, LAST ANNOTATION UPDATE)
 DE TOLUATE 1,2-DIOXYGENASE ELECTRON TRANSFER COMPONENT (CONTAINS:
 DE FERREDOXIN AND FERREDOXIN--NAD(+) REDUCTASE (EC 1.18.1.3)).
 GN XYLZ.
 OS PSEUDOMONAS PUTIDA.
 OG PLASMID TOL PWWO.
 OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; AEROBIC RODS AND COCCI;
 OC PSEUDOMONADACEAE.
 RN [1]
 RP SEQUENCE FROM N.A.
 RM 92041666
 RA HARAYAMA S., REKIK M., BAIROCH A., NEIDLE E.L., ORNSTON L.N.;
 RL J. BACTERIOL. 173:7540-7548(1991).
 CC -!- FUNCTION: ELECTRON TRANSFER COMPONENT OF TOLUATE 1,2-DIOXYGENASE
 CC SYSTEM.
 CC -!- SUBUNIT: THE DIOXYGENASE COMPLEX IS COMPOSED OF AN HYDROXYLASE
 CC COMPONENT THAT CONSISTS OF TWO CHAINS (XYLX AND XYLY), AND AN
 CC ELECTRON TRANSFER COMPONENT (XYLZ).
 CC -!- CATALYTIC ACTIVITY: REDUCED FERREDOXIN + NAD(+) = OXIDIZED
 CC FERREDOXIN + NADH.
 CC -!- SIMILARITY: IN THE N-TERMINAL REGION WITH 2FE-2S FERREDOXINS, AND
 CC IN THE REST OF THE SEQUENCE WITH FERREDOXIN REDUCTASE.
 CC -!- SIMILARITY: BELONGS TO THE

DR EMBL; M64747; PPXYL.
 DR PIR; C41659; C41659.
 DR PIR; S23484; S23484.
 DR PROSITE; PS00197; 2FE2S_FERREDOXIN.
 KW AROMATIC HYDROCARBONS CATABOLISM; FLAVOPROTEIN; OXIDOREDUCTASE;
 KW FAD; NAD; IRON-SULFUR; PLASMID.
 FT DOMAIN 29 98 FERREDOXIN.
 FT DOMAIN 99 336 FERREDOXIN-REDUCTASE.
 FT METAL 40 40 IRON-SULFUR (2FE-2S) (BY SIMILARITY).
 FT METAL 45 45 IRON-SULFUR (2FE-2S) (BY SIMILARITY).
 FT METAL 48 48 IRON-SULFUR (2FE-2S) (BY SIMILARITY).
 FT METAL 81 81 IRON-SULFUR (2FE-2S) (BY SIMILARITY).
 SQ SEQUENCE 336 AA; 36220 MW; 575585 CN;

DB 7; Score 66; Match 50.0%; Predicted No. 2.78e+00;
 Matches 10; Conservative 4; Mismatches 5; Indels 1; Gaps 1;

Db 298 evdiylcgpppm-veavsqq 316
 :||::| | | || |::|
 Qy 3 DVDLFLTGTPDEYVEQVAQY 22

RESULT 7

ID DP3A_ECOLI STANDARD; PRT; 1160 AA.
 AC P10443;
 DT 01-MAR-1989 (REL. 10, CREATED)
 DT 01-MAR-1989 (REL. 10, LAST SEQUENCE UPDATE)
 DT 01-JUN-1994 (REL. 29, LAST ANNOTATION UPDATE)
 DE DNA POLYMERASE III, ALPHA CHAIN (EC 2.7.7.7).
 GN DNAE OR POLC.
 OS ESCHERICHIA COLI.
 OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; FACULTATIVELY ANAEROBIC RODS;
 OC ENTEROBACTERIACEAE.
 RN [1]
 RP SEQUENCE FROM N.A.
 RM 88058791
 RA TOMASIEWICZ H.G., MCHENRY C.S.;
 RL J. BACTERIOL. 169:5735-5744(1987).
 RN [2]
 RP SEQUENCE OF 1070-1160 FROM N.A.
 RM 93123150
 RA LI S.J., CRONAN J.E. JR.;
 RL J. BACTERIOL. 175:332-340(1993).
 RN [3]
 RP REVIEW.
 RM 92246902
 RA O'DONNELL M.;
 RL BIOESSAYS 14:105-111(1992).
 RN [4]
 RP MUTAGENESIS.
 RM 93387658
 RA FIJALKOWSKA I.J., SCHAAPER R.M.;
 RL GENETICS 134:1039-1044(1993).
 CC -!- FUNCTION: DNA POLYMERASE III IS A COMPLEX, MULTICHAIN ENZYME
 CC RESPONSIBLE FOR MOST OF THE REPLICATIVE SYNTHESIS IN BACTERIA.
 CC THIS DNA POLYMERASE ALSO EXHIBITS 3' TO 5' EXONUCLEASE ACTIVITY.
 CC THE ALPHA CHAIN IS THE DNA POLYMERASE.
 CC -!- CATALYTIC ACTIVITY: N DEOXYNUCLEOSIDE TRIPHOSPHATE =
 CC N PYROPHOSPHATE + DNA(N).
 CC -!- SUBUNIT: CONTAINS A CORE (COMPOSED OF ALPHA, EPSILON, AND THETA
 CC CHAINS) THAT ASSOCIATES WITH A TAU SUBUNIT WHICH ALLOW THE CORE
 CC DIMERIZATION TO FORM THE POLIII' COMPLEX. POLIII' ASSOCIATES WITH
 CC THE GAMMA COMPLEX (COMPOSED OF CHAINS GAMMA, DELTA, DELTA', PSI,
 CC AND CHI) AND WITH THE BETA CHAIN. THE FINAL COMPOSITION OF THE
 CC COMPLEX IS: (ALPHA,EPSILON,THETA)[2]-TAU[2]-(GAMMA,DELTA,DELTA',
 CC PSI,CHI)[2]-BETA[2].

DR EMBL; M19334; ECLPXA.
DR EMBL; S52931; S52931.
DR PIR; C28390; DJEC3A.
DR PIR; A40637; A40637.
DR ECGENE; EG10238; DNAE.
KW DNA-DIRECTED DNA POLYMERASE; DNA REPLICATION.
SQ SEQUENCE 1160 AA; 129904 MW; 6441060 CN;

DB 2; Score 66; Match 35.0%; Predicted No. 2.78e+00;
Matches 7; Conservative 8; Mismatches 4; Indels 1; Gaps 1;

Db 957 lglyltghpinqylkeiery 976
: |||| | ::|: :: :|
Qy 4 VDLFLTGTP-DEYVEQVAQY 22

RESULT 8

ID DD01_HUMAN STANDARD; PRT: 1003 AA.
AC Q02218;
DT 01-JUL-1993 (REL. 26, CREATED)
DT 01-JUL-1993 (REL. 26, LAST SEQUENCE UPDATE)
DT 01-OCT-1993 (REL. 27, LAST ANNOTATION UPDATE)
DE 2-OXOGLUTARATE DEHYDROGENASE E1 COMPONENT PRECURSOR (EC 1.2.4.2)
DE (ALPHA-KETOGLUTARATE DEHYDROGENASE).
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; PRIMATES.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LIVER;
RM 92179301
RA KOIKE K., URATA Y., GOTO S.;
RL PROC. NATL. ACAD. SCI. U.S.A. 89:1963-1967(1992).
CC -!- FUNCTION: THE 2-OXOGLUTARATE DEHYDROGENASE COMPLEX CATALYZES THE
CC OVERALL CONVERSION OF 2-OXOGLUTARATE TO SUCCINYL-COA & CO(2). IT
CC CONTAINS MULTIPLE COPIES OF 3 ENZYMIC COMPONENTS: 2-OXOGLUTARATE
CC DEHYDROGENASE (E1), DIHYDROLIPOAMIDE SUCCINYLTRANSFERASE (E2) AND
CC LIPOAMIDE DEHYDROGENASE (E3).
CC -!- CATALYTIC ACTIVITY: 2-OXOGLUTARATE + LIPOAMIDE = S-SUCCINYL-
CC DIHYDROLIPOAMIDE + CO(2).
CC -!- COFACTOR: THIAMINE PYROPHOSPHATE.
CC -!- ENZYME REGULATION: CATABOLITE REPRESSED.
CC -!- SUBCELLULAR LOCATION: MITOCHONDRIAL MATRIX.
DR EMBL; D10523; HS20GDH.
DR PIR; A38234; A38234.
DR MIM; 203740; 11TH EDITION.
KW GLYCOLYSIS; OXIDOREDUCTASE; FLAVOPROTEIN; THIAMINE PYROPHOSPHATE;
KW MITOCHONDRION; TRANSIT PEPTIDE.
FT TRANSIT 1 40 MITOCHONDRION.
FT CHAIN 41 1003 ALPHA-KETOGLUTARATE DEHYDROGENASE.
SQ SEQUENCE 1003 AA; 113239 MW; 5162887 CN;

DB 5; Score 66; Match 46.7%; Predicted No. 2.78e+00;
Matches 7; Conservative 6; Mismatches 2; Indels 0; Gaps 0;

Db 47 epflsgtssnyveem 61
: ||: ||: : ||| :|
Qy 5 DLFLTGTPDEYVEQV 19

RESULT 9

ID DP3A_SALTY STANDARD; PRT: 1160 AA.
AC P14567;
DT 01-JAN-1990 (REL. 13, CREATED)
DT 01-JAN-1990 (REL. 13, LAST SEQUENCE UPDATE)
DT 01-JAN-1990 (REL. 13, LAST ANNOTATION UPDATE)

DE DNA POLYMERASE III, ALPHA CHAIN (EC 2.7.7.7).
 GN DNAE OR POLC.
 OS SALMONELLA TYPHIMURIUM.
 OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; FACULTATIVELY ANAEROBIC RODS;
 OC ENTEROBACTERIACEAE.
 RN [1]
 RP SEQUENCE FROM N.A.
 RM 90008797
 RA LANCY E.D., LIFSICS M.R., MUNSON P., MAURER R.;
 RL J. BACTERIOL. 171:5581-5586(1989).
 CC -!- FUNCTION: DNA POLYMERASE III IS A COMPLEX, MULTICHAIN ENZYME
 CC RESPONSIBLE FOR MOST OF THE REPLICATIVE SYNTHESIS IN BACTERIA.
 CC THIS DNA POLYMERASE ALSO EXHIBITS 3' TO 5' EXONUCLEASE ACTIVITY.
 CC THE ALPHA CHAIN IS THE DNA POLYMERASE.
 CC -!- CATALYTIC ACTIVITY: N DEOXYNUCLEOSIDE TRIPHOSPHATE =
 CC N PYROPHOSPHATE + DNA(N).
 CC -!- SUBUNIT: CONTAINS A CORE (COMPOSED OF ALPHA, EPSILON, AND THETA
 CC CHAINS) THAT ASSOCIATES WITH A TAU SUBUNIT WHICH ALLOW THE CORE
 CC DIMERIZATION TO FORM THE POLIII' COMPLEX. POLIII' ASSOCIATES WITH
 CC THE GAMMA COMPLEX (COMPOSED OF CHAINS GAMMA, DELTA, DELTA', PSI,
 CC AND CHI) AND WITH THE BETA CHAIN. THE FINAL COMPOSITION OF THE
 CC COMPLEX IS: (ALPHA,EPSILON,THETA)[2]-TAU[2]-(GAMMA,DELTA,DELTA',
 CC PSI,CHI)[2]-BETA[4].
 DR EMBL: M29701; STDNAE.
 DR EMBL: M26046; STPOL3A.
 DR PIR: A45915; A45915.
 KW DNA-DIRECTED DNA POLYMERASE; DNA REPLICATION.
 SQ SEQUENCE 1160 AA; 130118 MW; 6471246 CN;

DB 2; Score 66; Match 35.0%; Predicted No. 2.78e+00;
 Matches 7; Conservative 8; Mismatches 4; Indels 1; Gaps 1;

Db 957 lglyltghpinqylkeiery 976
 : | : | | | : : | : : : |
 Qy 4 VDLFLTGTP-DEYVEQVAQY 22

RESULT 10
 ID AGRI_CHICK STANDARD; PRT; 1955 AA.
 AC P31696;
 DT 01-JUL-1993 (REL. 26, CREATED)
 DT 01-JUL-1993 (REL. 26, LAST SEQUENCE UPDATE)
 DT 01-JUL-1993 (REL. 26, LAST ANNOTATION UPDATE)
 DE AGRIN PRECURSOR.
 OS GALLUS GALLUS (CHICKEN).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; AVES; NEOGNATHAE;
 OC GALLIFORMES.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=BRAIN;
 RM 92232297
 RA TSIM K.W.K., RUEGG M.A., ESCHER G., KROEGER S., MCMAHAN U.J.;
 RL NEURON 8:677-689(1992).
 RN [2]
 RP ALTERNATIVE SPLICING.
 RM 92232298
 RA RUEGG M.A., TSIM K.W.K., HORTON S.E., KROEGER S., ESCHER G.,
 RA GENSCH E.M., MCMAHAN U.J.;
 RL NEURON 8:691-699(1992).
 CC -!- FUNCTION: COMPONENT OF THE BASAL LAMINA THAT CAUSES THE
 CC AGGREGATION OF ACETYLCHOLINE RECEPTORS AND ACETYLCHOLINE-ESTERASE
 CC ON THE SURFACE OF MUSCLE FIBERS OF THE NEUROMUSCULAR JUNCTION.
 CC -!- SUBCELLULAR LOCATION: SYNAPTIC BASAL LAMINA AT THE NEUROMUSCULAR
 CC JUNCTION.
 CC -!- ALTERNATIVE PRODUCTS: AT LEAST THREE DIFFERENT FORMS ARISE BY
 CC ALTERNATIVE SPLICING. THE LARGEST, A FULL ACETYLCHOLINE RECEPTOR

CC CLUSTERING ACTIVITY.

DR EMBL; M94271; GGAGRIN.

DR EMBL; M97371; GGAGRPR1A.

DR EMBL; M97372; GGAGRPR2A.

DR PIR; JH0591; AGCH.

DR PROSITE; PS00022; EGF.

KW GLYCOPROTEIN; EGF-LIKE DOMAIN; REPEAT; ALTERNATIVE SPLICING; SIGNAL.

FT	SIGNAL	1	38	POTENTIAL.
FT	CHAIN	39	1955	AGRIN.
FT	DOMAIN	54	126	KAZAL-LIKE.
FT	DOMAIN	130	201	KAZAL-LIKE.
FT	DOMAIN	202	273	KAZAL-LIKE.
FT	DOMAIN	276	344	KAZAL-LIKE.
FT	DOMAIN	350	418	KAZAL-LIKE.
FT	DOMAIN	419	483	KAZAL-LIKE.
FT	DOMAIN	484	548	KAZAL-LIKE.
FT	DOMAIN	551	633	KAZAL-LIKE.
FT	SIMILAR	687	793	LAMININ DOMAIN III.
FT	REPEAT	688	739	
FT	REPEAT	742	786	
FT	DOMAIN	781	851	KAZAL-LIKE.
FT	DOMAIN	1233	1751	4 X EGF-TYPE REPEATS.
FT	REPEAT	1233	1264	EGF-LIKE 1.
FT	REPEAT	1450	1482	EGF-LIKE 2.
FT	REPEAT	1489	1521	EGF-LIKE 3.
FT	REPEAT	1718	1751	EGF-LIKE 4.
FT	DOMAIN	856	995	SER/THR-RICH.
FT	DOMAIN	1150	1219	SER/THR-RICH.
FT	VARSPPLIC	1648	1651	MISSING (IN AGRIN-RELATED PROTEIN 2).
FT	VARSPPLIC	1784	1793	MISSING (IN AGRIN-RELATED PROTEINS 1 AND 2).
FT				
FT	DISULFID	86	105	POTENTIAL.
FT	DISULFID	94	126	POTENTIAL.
FT	DISULFID	160	180	POTENTIAL.
FT	DISULFID	169	201	POTENTIAL.
FT	DISULFID	233	252	POTENTIAL.
FT	DISULFID	241	273	POTENTIAL.
FT	DISULFID	304	323	POTENTIAL.
FT	DISULFID	312	344	POTENTIAL.
FT	DISULFID	378	397	POTENTIAL.
FT	DISULFID	386	418	POTENTIAL.
FT	DISULFID	443	462	POTENTIAL.
FT	DISULFID	451	483	POTENTIAL.
FT	DISULFID	507	527	POTENTIAL.
FT	DISULFID	516	548	POTENTIAL.
FT	DISULFID	592	612	POTENTIAL.
FT	DISULFID	601	633	POTENTIAL.
FT	DISULFID	810	830	POTENTIAL.
FT	DISULFID	819	851	POTENTIAL.
FT	DISULFID	1233	1244	POTENTIAL.
FT	DISULFID	1238	1253	POTENTIAL.
FT	DISULFID	1255	1264	POTENTIAL.
FT	DISULFID	1450	1461	POTENTIAL.
FT	DISULFID	1455	1471	POTENTIAL.
FT	DISULFID	1473	1482	POTENTIAL.
FT	DISULFID	1489	1500	POTENTIAL.
FT	DISULFID	1494	1510	POTENTIAL.
FT	DISULFID	1512	1521	POTENTIAL.
FT	CARBOHYD	390	390	POTENTIAL.
FT	CARBOHYD	659	659	POTENTIAL.
FT	CARBOHYD	764	764	POTENTIAL.
FT	CARBOHYD	814	814	POTENTIAL.
SO	SEQUENCE	1955 AA;	211411 MW;	17426209 CN;

Db 1387 dtdlfvvgapedqnavvaertaatv 1411
 | |||: ||: : ||: | |
 Qy 3 DVDLFLTGPDEYVEQVAQYKALPV 27

RESULT 11

ID ANX1_MOUSE STANDARD; PRT; 345 AA.
 AC P10107;
 DT 01-MAR-1989 (REL. 10, CREATED)
 DT 01-MAR-1989 (REL. 10, LAST SEQUENCE UPDATE)
 DT 01-OCT-1994 (REL. 30, LAST ANNOTATION UPDATE)
 DE ANNEXIN I (LIPOCORTIN I) (CALPACTIN II) (CHROMOBINDIN 9) (P35)
 DE (PHOSPHOLIPASE A2 INHIBITORY PROTEIN).
 GN LPC-1.
 OS MUS MUSCULUS (MOUSE).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
 OC EUTHERIA; RODENTIA.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=DS;
 RM 89098333
 RA SAKATA T., IWAGAMI S., TSURUTA Y., SUZUKI R., HOJO K., SATO K.,
 RA TERAOKA H.;
 RL NUCLEIC ACIDS RES. 16:11818-11818(1988).
 RN [2]
 RP SEQUENCE OF 5-345 FROM N.A.
 RM 89165848
 RA PHILIPPS C., ROSE-JOHN S., RINCKE G., FUERSTENBERGER G., MARKS F.;
 RL BIOCHEM. BIOPHYS. RES. COMMUN. 159:155-162(1989).
 CC -!- FUNCTION: THIS PROTEIN REGULATES PHOSPHOLIPASE A2 ACTIVITY. IT
 CC SEEMS TO BIND FROM TWO TO FOUR CALCIUM IONS WITH HIGH AFFINITY.
 CC -!- PTM: PHOSPHORYLATION OF ANNEXIN 1 RESULTS IN LOSS OF ITS
 CC INHIBITORY ACTIVITY.
 CC -!- DOMAIN: CONTAINS FOUR HOMOLOGOUS REPEATS WITH A CONSENSUS
 CC SEQUENCE COMMON TO ALL ANNEXIN PROTEINS. A PAIR OF THESE REPEATS
 CC MAY FORM ONE BINDING SITE FOR CALCIUM AND PHOSPHOLIPID.
 CC -!- SIMILARITY: TO OTHER PROTEINS OF THE ANNEXIN FAMILY.
 DR EMBL; X07486; MMLCIR.
 DR EMBL; M24554; MMLCI.
 DR PIR; S02181; LUMS1.
 DR PROSITE; PS00223; ANNEXIN.
 KW ANNEXIN; CALCIUM/PHOSPHOLIPID-BINDING; REPEAT;
 KW PHOSPHOLIPASE A2 INHIBITOR; PHOSPHORYLATION.
 FT INIT_MET 0 0
 FT REPEAT 50 110 ANNEXIN.
 FT REPEAT 122 182 ANNEXIN.
 FT REPEAT 206 266 ANNEXIN.
 FT REPEAT 281 341 ANNEXIN.
 FT MOD_RES 20 20 PHOSPHORYLATION (BY TYR-KINASES).
 FT CONFLICT 77 78 QQ -> PR (IN REF. 2).
 FT CONFLICT 221 221 T -> H (IN REF. 2).
 FT CONFLICT 273 273 T -> H (IN REF. 2).
 SQ SEQUENCE 345 AA; 38603 MW; 569605 CN;

DB 1; Score 65; Match 45.0%; Predicted No. 3.97e+00;
 Matches 9; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

Db 12 flenqeqyqvayksykggp 31
 || :|||: | ||: |
 Qy 7 FLTGTGPDEYVEQVAQYKALP 26

RESULT 12

ID RYNC_RABIT STANDARD; PRT; 4969 AA.
 AC P10107;

DT 01-JUL-1993 (REL. 26, CREATED)
 DT 01-JUL-1993 (REL. 26, LAST SEQUENCE UPDATE)
 DT 01-JUL-1993 (REL. 26, LAST ANNOTATION UPDATE)
 DE RYANODINE RECEPTOR, CARDIAC MUSCLE.
 OS ORYCTOLAGUS CUNICULUS (RABBIT).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
 OC EUTHERIA; LAGOMORPHA.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=CARDIAC MUSCLE;
 RM 90337947
 RA OTSU K., WILLARD H.F., KHANNA V.K., ZORZATO F., GREEN N.M.,
 RA MACLENNAN D.H.;
 RL J. BIOL. CHEM. 265:13472-13483(1990).
 RN [2]
 RP PHOSPHORYLATION OF SER-2809.
 RM 91250425
 RA WITCHER D.R., KOVACS R.J., SCHULMAN H., CEFALI D.C., JONES L.R.;
 RL J. BIOL. CHEM. 266:11144-11152(1991).
 CC -!- FUNCTION: COMMUNICATION BETWEEN TRANSVERSE-TUBULES AND SARCOPLAMIC
 CC RETICULUM. CONTRACTION OF CARDIAC MUSCLE IS TRIGGERED BY RELEASE
 CC OF CA++ FROM SR FOLLOWING DEPOLARIZATION OF T-TUBULES.
 CC -!- THE CALCIUM RELEASE CHANNEL IS MODULATED BY CA++, MG++, ATP, AND
 CC CALMODULIN.
 CC -!- THE CALCIUM RELEASE CHANNEL ACTIVITY RESIDES IN THE C-TERMINAL
 CC REGION WHILE THE REMAINING PART OF THE PROTEIN CONSISTUTES THE
 CC 'FOOT' STRUCTURE SPANNING THE JUNCTIONAL GAP BETWEEN THE SR AND
 CC THE T-TUBULE. IT IS POSSIBLE THAT THE FOOT STRUCTURE INTERACTS
 CC WITH THE CYTOPLASMIC REGION OF THE DIHYDROPYRIDINE RECEPTOR.
 CC -!- RYANODINE IS AN ALKALOID THAT BINDS TO THE CA-RELEASE CHANNEL IN
 CC JUNCTIONAL SR AND MODULATES ITS ACTIVITY.
 CC -!- SUBUNIT: HOMOTETRAMER (POTENTIAL).
 CC -!- TISSUE SPECIFICITY: HEART AND BRAIN.
 CC -!- SIMILARITY: LOCAL & LOW WITH THE NICOTINIC ACETYLCHOLINE RECEPTOR
 CC (N-ACHR) SUBUNITS.
 DR EMBL: M59743; OCCA2RE.
 DR PIR: A37113; A37113.
 KW RECEPTOR; TRANSMEMBRANE; IONIC CHANNEL; CALCIUM CHANNEL; REPEAT;
 KW PHOSPHORYLATION; GLYCOPROTEIN.
 FT DOMAIN 1 3090 CYTOPLASMIC.
 FT TRANSMEM 3091 3110 M' (POTENTIAL).
 FT TRANSMEM 3154 3172 M'' (POTENTIAL).
 FT TRANSMEM 3941 3960 M1 (POTENTIAL).
 FT TRANSMEM 3979 3996 M2 (POTENTIAL).
 FT TRANSMEM 4234 4257 M3 (POTENTIAL).
 FT TRANSMEM 4295 4315 M4 (POTENTIAL).
 FT TRANSMEM 4501 4521 M5 (POTENTIAL).
 FT TRANSMEM 4580 4602 M6 (POTENTIAL).
 FT TRANSMEM 4722 4752 M7 (POTENTIAL).
 FT TRANSMEM 4770 4788 M8 (POTENTIAL).
 FT TRANSMEM 4812 4829 M9 (POTENTIAL).
 FT TRANSMEM 4847 4869 M10 (POTENTIAL).
 FT DOMAIN 853 2926 4 X APPROXIMATE REPEATS.
 FT REPEAT 853 966 1.
 FT REPEAT 967 1080 2.
 FT REPEAT 2693 2811 3.
 FT REPEAT 2813 2926 4.
 FT BINDING 2619 3016 MODULATOR (POTENTIAL).
 FT BINDING 2775 2807 CALMODULIN (POTENTIAL).
 FT BINDING 2877 2898 CALMODULIN (POTENTIAL).
 FT BINDING 2998 3016 CALMODULIN (POTENTIAL).
 FT MOD_RES 2809 2809 PHOSPHORYLATION (BY CAM-KINASE).
 FT CARBOHYD 198 198 POTENTIAL.
 FT CARBOHYD 404 404 POTENTIAL.
 FT CARBOHYD 1636 1636 POTENTIAL.
 FT CARBOHYD 1636 1636 POTENTIAL.

```

FT CARBOHYD 2224 2224 POTENTIAL.
FT CARBOHYD 2803 2803 POTENTIAL.
FT CARBOHYD 2831 2831 POTENTIAL.
FT CARBOHYD 3096 3096 POTENTIAL.
FT CARBOHYD 4105 4105 POTENTIAL.
FT CARBOHYD 4796 4796 POTENTIAL.
SQ SEQUENCE 4969 AA; 565060 MW; 24964830 CN;

```

DB 6; Score 65; Match 38.1%; Predicted No. 3.97e+00;
 Matches 8; Conservative 3; Mismatches 10; Indels 0; Gaps 0;

```

Db 3616 ravnlfqgyekswieteehy 3636
   | | : | | | | : | : |
Qy 2 RDVDLFLTGTPEYVEQVAQY 22

```

RESULT 13

```

ID GLSF_ANTSP STANDARD; PRT: 1536 AA.
AC Q06434;
DT 01-JUN-1994 (REL. 29, CREATED)
DT 01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)
DT 01-OCT-1994 (REL. 30, LAST ANNOTATION UPDATE)
DE FERREDOXIN-DEPENDENT GLUTAMATE SYNTHASE (EC 1.4.7.1) (FD-GOGAT).
GN GLTB OR GLSF.
OS ANTITHAMNION SP.
OG CHLOROPLAST.
OC EUKARYOTA; PLANTA; PHYCOPHYTA; RHODOPHYTA (RED ALGAE).
RN [1]
RP SEQUENCE FROM N.A.
RM 94033299
RA VALENTIN K.U., KOSTRZEWA M., ZETSCHE K.;
RL PLANT MOL. BIOL. 23:77-85(1993).
CC -!- CATALYTIC ACTIVITY: 2 L-GLUTAMATE + 2 OXIDIZED FERREDOXIN =
CC L-GLUTAMINE + 2-OXOGLUTARATE + 2 REDUCED FERREDOXIN.
CC -!- COFACTOR: IRON-SULFUR; FAD AND FMN FLAVOPROTEIN.
CC -!- PATHWAY: GLUTAMINE SYNTHETASE/GOGAT PATHWAY WHICH IS INVOLVED
CC IN THE ASSIMILATION OF AMMONIA.
CC -!- SUBUNIT: MONOMER.
CC -!- SUBCELLULAR LOCATION: CHLOROPLAST STROMA.
CC -!- SIMILARITY: TO OTHER GLUTAMATE SYNTHASES.
DR EMBL; Z21705; CHASGLTB.
DR PIR; S31911; S31911.
DR PIR; S39510; S39510.
KW OXIDOREDUCTASE; IRON-SULFUR; FLAVOPROTEIN; FAD; FMN; CHLOROPLAST;
KW GLUTAMATE BIOSYNTHESIS.
FT NP_BIND 1105 1162 FMN (BY SIMILARITY).
SQ SEQUENCE 1536 AA; 171111 MW; 12053975 CN;

```

DB 3; Score 65; Match 33.3%; Predicted No. 3.97e+00;
 Matches 5; Conservative 6; Mismatches 4; Indels 0; Gaps 0;

```

Db 1340 kgihtlylkgeandyv 1354
   : : | : | | : : : | |
Qy 2 RDVDLFLTGTPEYV 16

```

RESULT 14

```

ID VG67_HSVI1 STANDARD; PRT: 1556 AA.
AC Q00107;
DT 01-DEC-1992 (REL. 24, CREATED)
DT 01-DEC-1992 (REL. 24, LAST SEQUENCE UPDATE)
DT 01-DEC-1992 (REL. 24, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL GENE 67 PROTEIN.
GN 67.
OS ICTALURID HERPESVIRUS 1 (CHANNEL CATFISH VIRUS) (CCV).
OC HERPESVIRIDAE; GENE 67 PROTEIN; VIRUS; HERPESVIRIDAE; ALPHAHERPESVIRINAE

```

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AUBURN 1;
RM 92087490
RA DAVISON A.J.;
RL VIROLOGY 186:9-14(1992).
DR EMBL; M75136; HECHCCOMG.
DR PIR; D36793; D36793.
KW HYPOTHETICAL PROTEIN.
SQ SEQUENCE 1556 AA; 173577 MW; 12789685 CN;

DB 7; Score 64; Match 34.6%; Predicted No. 5.64e+00;
Matches 9; Conservative 6; Mismatches 11; Indels 0; Gaps 0;

Db 1201 ravesfmlrdparyivelapegs1pv 1226
| | : | : : : :
Qy 2 RDVDLFLTGTPEYVEQVAGYKALPV 27

RESULT 15

ID SUI5_RAT STANDARD; PRT: 917 AA.
AC P23739;
DT 01-NOV-1991 (REL. 20, CREATED)
DT 01-AUG-1992 (REL. 23, LAST SEQUENCE UPDATE)
DT 01-FEB-1994 (REL. 28, LAST ANNOTATION UPDATE)
DE SUCRASE-ISOMALTASE, INTESTINAL (EC 3.2.1.48) / (EC 3.2.1.10)
DE (FRAGMENTS).
OS RATTUS NORVEGICUS (RAT).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; RODENTIA.
RN [1]
RP SEQUENCE OF 1-275 FROM N.A.
RC STRAIN=FISHER 344; TISSUE=INTESTINE;
RM 91097578
RA TRABER P.G.;
RL BIOCHEM. BIOPHYS. RES. COMMUN. 173:765-773(1990).
RN [2]
RP SEQUENCE OF 276-917 FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=DUODENUM;
RM 90381315
RA BROYART J.-P., HUGOT J.-P., PERRET C., PORTEU A.;
RL BIOCHIM. BIOPHYS. ACTA 1087:61-67(1990).
CC -!- FUNCTION: PLAYS AN IMPORTANT ROLE IN THE FINAL STAGE OF
CC CARBOHYDRATE DIGESTION.
CC -!- CATALYTIC ACTIVITY: HYDROLYSIS OF SUCROSE AND MALTOSE BY AN
CC ALPHA-D-GLUCOSIDASE-TYPE ACTION.
CC -!- CATALYTIC ACTIVITY: HYDROLYSIS OF 1,6-ALPHA-D-GLUCOSIDIC LINKAGES
CC IN ISOMALTOSE AND DEXTRINS PRODUCED FROM STARCH AND GLYCOGEN BY
CC ALPHA-AMYLASE.
CC -!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. BRUSH BORDER.
CC -!- PTM: THE PRECURSOR IS PROTEOLYTICALLY CLEAVED WHEN EXPOSED TO
CC PANCREATIC PROTEASES IN THE INTESTINAL LUMEN.
CC -!- SUBUNIT: THE RESULTING SUCRASE AND ISOMALTASE SUBUNITS STAY
CC ASSOCIATED WITH ONE ANOTHER IN A COMPLEX BY NON-COVALENT LINKAGES.
CC -!- THERE IS A HIGH DEGREE OF HOMOLOGY BETWEEN THE ISOMALTASE AND
CC SUCRASE PORTIONS (41 % OF AMINO ACID IDENTITY) INDICATING THAT
CC THIS PROTEIN IS EVOLVED BY PARTIAL GENE DUPLICATION.
CC -!- SIMILARITY: BELONGS TO FAMILY 31 OF GLYCOSYL HYDROLASES.
DR EMBL; M62889; RRS1.
DR PIR; S11386; S11386.
DR PROSITE; PS00129; GLYCOSYL_HYDROL_F31_1.
DR PROSITE; PS00707; GLYCOSYL_HYDROL_F31_2.
KW MULTIFUNCTIONAL ENZYME; INTESTINE; TRANSMEMBRANE; GLYCOPROTEIN;
KW HYDROLASE; GLYCOSIDASE; DUPLICATION.
FT NON_TER 1 1
FT

```

FT  NON_CONS      275    276
FT  DOMAIN        <276    557    ISOMALTASE.
FT  DOMAIN        558    >917    SUCRASE.
FT  CARBOHYD      23     23     POTENTIAL.
FT  CARBOHYD      302    302    POTENTIAL.
FT  CARBOHYD      309    309    POTENTIAL.
FT  CARBOHYD      411    411    POTENTIAL.
FT  CARBOHYD      454    454    POTENTIAL.
FT  CARBOHYD      784    784    POTENTIAL.
FT  CARBOHYD      852    852    POTENTIAL.
FT  CARBOHYD      889    889    POTENTIAL.
FT  CARBOHYD      903    903    POTENTIAL.
FT  NON_TER       917    917
SQ  SEQUENCE      917 AA; 105272 MW; 4481141 CN;

```

DB 6; Score 64; Match 33.3%; Predicted No. 5.64e+00;
 Matches 7; Conservative 8; Mismatches 6; Indels 0; Gaps 0;

Db 249 iflgdtpeqvqqyqefngrp 269
 :|| ||: ||: ||: ||: |
 Qy 6 LFLTGTPDEYVEQVAQYKALP 26

Search completed: Fri Mar 24 07:41:16 1995
 Job time : 13 secs.

MASPAR (TM)

Release 2.0 John F. Collins & S. S. Sturrock, Biocomputing Research Unit.
 Copyright (c) 1993, 1994 by University of Edinburgh, U.K.
 Distribution rights by IntelliGenetics, Inc.

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Mar 24 07:44:01 1995; MasPar time 2.86 Seconds
 57.341 Million cell updates/sec

Tabular output not generated.

Title: >US-08-300-510-2
 Description: (1:27) from US08300510.pep
 Perfect Score: 184
 Sequence: 1 KALPVVLENARILKNCVDKMTTEEDKE 27

Scoring table: PAM 150
 Gap 14

Searched: 50375 seqs, 6065180 residues

Database: a-geneseq
 1 a-gen1
 2 a-gen2
 3 a-gen3
 4 a-gen4
 5 a-gen5
 6 a-gen6
 7 a-gen7
 8 a-gen8

9 a-gen9
10 a-gen10

Statistics: Mean 21.324; Variance 75.358; scale 0.283

Predicted No. is the number of results expected by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	184	100.0	88	8	R41984	Human T cell reactive	4.68e-12
2	184	100.0	92	8	R41983	Human T cell reactive	4.68e-12
3	184	100.0	27	8	R41976	Human T cell reactive	4.68e-12
4	184	100.0	96	7	R36548	Recombitope YZX.	4.68e-12
5	184	100.0	94	3	R12119	TRFP chain 1 with lea	4.68e-12
6	184	100.0	92	7	R36539	TRFP chain 1 (with Le	4.68e-12
7	184	100.0	27	7	R36543	Peptide Y.	4.68e-12
8	184	100.0	88	7	R36540	TRFP chain 1 (with Le	4.68e-12
9	184	100.0	96	3	R12120	TRFP chain 1 with lea	4.68e-12
10	184	100.0	96	5	R27368	TRFP Chain #1 with C1	4.68e-12
11	184	100.0	94	5	R27367	TRFP Chain #1 with C1	4.68e-12
12	68	37.0	333	1	P91948	Pig H4 isoenzyme.	6.74e+00
13	60	32.6	1513	8	R43253	p190 protein.	3.60e+01
14	60	32.6	520	1	R04571	ORF3 product from the	3.60e+01
15	59	32.1	765	1	P92275	Human topoisomerase I	4.41e+01
16	59	32.1	631	2	P70674	Mx protein.	4.41e+01
17	58	31.5	990	7	R39343	EpiB protein.	5.40e+01
18	58	31.5	844	5	R25671	Mouse vav proto oncog	5.40e+01
19	57	31.0	166	4	P50193	Modified human interf	6.61e+01
20	57	31.0	166	4	P50206	Modified human interf	6.61e+01
21	57	31.0	166	4	P50196	Modified human interf	6.61e+01
22	57	31.0	166	4	P50194	Modified human interf	6.61e+01
23	57	31.0	165	4	P50207	Modified human interf	6.61e+01
24	56	30.4	1146	3	R15156	Abelson Related Gene,	8.08e+01
25	55	29.9	719	9	R49507	Human LIF-R clone 65.	9.86e+01
26	55	29.9	719	5	R25069	mLIF-R.	9.86e+01
27	54	29.3	734	4	R20210	Diacylglycerol kinas	1.20e+02
28	54	29.3	558	9	R44140	Murine FACC encoded b	1.20e+02
29	54	29.3	95	4	R20983	Sequence encoded by a	1.20e+02
30	54	29.3	591	9	R51270	Murine FACC encoded b	1.20e+02
31	54	29.3	836	8	R41869	Human OSF-2.	1.20e+02
32	53	28.8	1284	1	P81187	Sequence encoded by a	1.46e+02
33	53	28.8	738	3	R13251	PECAM-1.	1.46e+02
34	53	28.8	354	2	R11789	Histo-blood gp. A gly	1.46e+02
35	53	28.8	354	2	R11792	Histo-blood gp. B gly	1.46e+02
36	53	28.8	354	2	R11790	Histo-blood gp. B gly	1.46e+02
37	53	28.8	672	2	R03725	Human placenta-derive	1.46e+02
38	53	28.8	353	2	R11317	Histo-blood gp. A gly	1.46e+02
39	53	28.8	1500	6	R30636	hCPSI.	1.46e+02
40	53	28.8	241	4	R23007	Protein transcribed f	1.46e+02
41	53	28.8	493	7	R33890	mec-4 protein.	1.46e+02
42	53	28.8	493	4	R20107	Mec-4 gene product.	1.46e+02
43	53	28.8	786	2	R03726	Human placenta-derive	1.46e+02
44	53	28.8	143	3	P50063	Soybean heat shock pr	1.46e+02
45	53	28.8	1048	4	R20637	Human Cytomegalovirus	1.46e+02

ALIGNMENTS

RESULT 1

ID R41984 standard; Protein; 88 AA.

AC 000000

DT 21-APR-1994 (first entry)
 DE Human T cell reactive feline protein B chain 1.
 KW Human; T cell; reactive; feline; protein; immune response; antigen;
 KW tolerance; mammal; Dermatophagoides; Felis; Ambrosia; Lolium; Canis;
 KW Cryptomeria; Alternaria; Alder; Betula; Quercus; Olea; Artemesia;
 KW Plantago; Parietaria; Blattella; Apis; Periplaneta; autoantigen.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT Peptide 1..17
 FT /note= "Signal peptide"
 FT Protein 18..88
 FT /note= "Mature protein"
 PN WD9319178-A.
 PD 30-SEP-1993.
 PF 25-MAR-1993; U02462.
 PR 25-MAR-1992; US-857311.
 PR 15-MAY-1992; US-884718.
 PR 15-JAN-1993; US-006116.
 PA (IMMU-) IMMUNOLOGIC PHARM CORP.
 PI Briner TJ, Garman RD, Geffer ML, Greenstein JL, Kuo M;
 PI Morville M;
 DR WPI; 93-320744/40.
 DR N-PSDB; Q49534.
 PT New peptide(s) for inducing tolerance - comprise one or more
 PT epitope(s) of an allergen administered subcutaneously, for
 PT treating sensitivity to cats, bees, etc.
 PS Disclosure; Fig 1; 107pp; English.
 CC The sequences given in R41983-84 represent chain 1 of human T cell
 CC reactive feline proteins (TRFP) A and B respectively. Peptides
 CC derived from TRFP may be used in a therapeutic composition which is
 CC useful in treating diseases which involve an immune response to a
 CC protein antigen. This composition may be used to induce tolerance
 CC in a mammal to Dermatophagoides, Felis, Ambrosia, Lolium, Cryptomeria,
 CC Alternaria, Alder, Betula, Quercus, Olea, Artemesia, Plantago,
 CC Parietaria, Canis, Blattella, Apis, Periplaneta and to autoantigens
 CC in humans.
 SQ Sequence 88 AA;

DB 8; Score 184; Match 100.0%; Predicted No. 4.68e-12;
 Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 47 kalpvvlenarilkncvdakmteedke 73
 |||||
 Qy 1 KALPVVLENARILKNCVDAKMTEEDKE 27

RESULT 2

ID R41983 standard; Protein; 92 AA.
 AC R41983;
 DT 21-APR-1994 (first entry)
 DE Human T cell reactive feline protein A chain 1.
 KW Human; T cell; reactive; feline; protein; immune response; antigen;
 KW tolerance; mammal; Dermatophagoides; Felis; Ambrosia; Lolium; Canis;
 KW Cryptomeria; Alternaria; Alder; Betula; Quercus; Olea; Artemesia;
 KW Plantago; Parietaria; Blattella; Apis; Periplaneta; autoantigen.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT Peptide 1..22
 FT /note= "Signal peptide"
 FT Protein 23..92
 FT /note= "Mature protein"
 PN WD9319178-A.
 PD 30-SEP-1993.
 PF 25-MAR-1993; U02462.
 PR 25-MAR-1992; US-857311.
 PR 15-MAY-1992; US-884718.

PR 15-JAN-1993; US-006116.
 PA (IMMU-) IMMUNOLOGIC PHARM CORP.
 PI Briner TJ, Garman RD, Gefter ML, Greenstein JL, Kuo M;
 PI Morville M;
 DR WPI; 93-320744/40.
 DR N-PSDB; Q49533.
 PT New peptide(s) for inducing tolerance - comprise one or more
 PT epitope(s) of an allergen administered subcutaneously, for
 PT treating sensitivity to cats, bees, etc.
 PS Disclosure; Fig 1; 107pp; English.
 CC The sequences given in R41983-84 represent chain 1 of human T cell
 CC reactive feline proteins (TRFP) A and B respectively. Peptides
 CC derived from TRFP may be used in a therapeutic composition which is
 CC useful in treating diseases which involve an immune response to a
 CC protein antigen. This composition may be used to induce tolerance
 CC in a mammal to Dermatophagoides, Felis, Ambrosia, Lolium, Cryptomeria,
 CC Alternaria, Alder, Betula, Quercus, Olea, Artemesia, Plantago,
 CC Parietaria, Canis, Blattella, Apis, Periplaneta and to autoantigens
 CC in humans.
 SQ Sequence 92 AA;

DB 8; Score 184; Match 100.0%; Predicted No. 4.68e-12;
 Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 51 kalpvvlenarilkncvdakmteedke 77
 ||||||||||||||||||||
 Qy 1 KALPVVLENARILKNCVDAKMTEEDKE 27

RESULT 3

ID R41976 standard; peptide; 27 AA.
 AC R41976;
 DT 21-APR-1994 (first entry)
 DE Human T cell reactive feline protein fragment Y.
 KW Human; T cell; reactive; feline; protein; immune response; antigen;
 KW tolerance; mammal; Dermatophagoides; Felis; Ambrosia; Lolium; Canis;
 KW Cryptomeria; Alternaria; Alder; Betula; Quercus; Olea; Artemesia;
 KW Plantago; Parietaria; Blattella; Apis; Periplaneta; autoantigen; ss.
 OS Homo sapiens.
 PN WD9319178-A.
 PD 30-SEP-1993.
 PF 25-MAR-1993; U02462.
 PR 25-MAR-1992; US-857311.
 PR 15-MAY-1992; US-884718.
 PR 15-JAN-1993; US-006116.
 PA (IMMU-) IMMUNOLOGIC PHARM CORP.
 PI Briner TJ, Garman RD, Gefter ML, Greenstein JL;
 PI Kuo M, Morville M;
 DR WPI; 93-320744/40.
 PT New peptide(s) for inducing tolerance - comprise one or more
 PT epitope(s) of an allergen administered subcutaneously, for
 PT treating sensitivity to cats, bees, etc.
 PS Claim 1; Fig 3; 107pp; English.
 CC The sequences given in R41975-82 are peptides derived from a human T
 CC cell reactive feline protein. These peptides are used in a
 CC therapeutic composition which is useful in treating diseases which
 CC involve an immune response to a protein antigen. This composition
 CC may be used to induce tolerance in a mammal to Dermatophagoides,
 CC Felis, Ambrosia, Lolium, Cryptomeria, Alternaria, Alder, Betula,
 CC Quercus, Olea, Artemesia, Plantago, Parietaria, Canis, Blattella,
 CC Apis, Periplaneta and to autoantigens in humans.
 SQ Sequence 27 AA;

DB 8; Score 184; Match 100.0%; Predicted No. 4.68e-12;
 Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 kalpvvlenarilkncvdakmteedke 27
|||||
Qy 1 KALPVVLENARILKNCVDAKMTEEDKE 27

RESULT 4

ID R36548 standard; Protein; 96 AA.
AC R36548;
DT 12-AUG-1993 (first entry)
DE Recombitope YZX.
KW Human T cell reactive feline protein; TRFP; epitope; recombitepo
KW sensitivity; Felis domesticus.
OS Synthetic.
FH Key Location/Qualifiers
FT Cleavage_site 14..15
FT /label= thrombin_cleavage_site
PN WD9308280-A.
PD 29-APR-1993.
PF 16-OCT-1992; U08694.
PR 16-OCT-1991; US-777859.
PR 13-DEC-1991; US-807529.
PA (IMMU-) IMMULOGIC PHARM CORP.
PI Bond JF, Garman RD, Kuo M, Morgenstern JP, Morville M;
PI Rogers BL;
DR WPI; 93-152473/18.
DR N-PSDB; Q41572.
PT Recombitope peptide having T-cell stimulating activity - for the
PT diagnosis and treatment of sensitivity to protein allergens,
PT auto:antigens and protein antigens
PS Disclosure; Fig 8; 73pp; English.
CC Preferred recombitepo peptides for treating sensitivity to Felis
CC domesticus are derived from the the genus Felis and comprise
CC regions selected from peptides X, Y, Z, A and B, of TRFP, and
CC modifications thereof, such as peptide C.
CC Oligonucleotides C, D, E, F, G, H and I are used in the
CC construction of recombitepo peptide YZX.
SQ Sequence 96 AA;

DB 7; Score 184; Match 100.0%; Predicted No. 4.68e-12;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 17 kalpvvlenarilkncvdakmteedke 43
|||||
Qy 1 KALPVVLENARILKNCVDAKMTEEDKE 27

RESULT 5

ID R12119 standard; Protein; 94 AA.
AC R12119;
DT 26-JUL-1991 (first entry)
DE TRFP chain 1 with leader A.
KW Human T cell reactive feline protein; cat allergens.
OS Felis catus.
FH Key Location/Qualifiers
FT Peptide 3..24
FT /label= Leader B
FT Protein 25..94
FT /label= TRFP Chain 1
PN WD9106571-A.
PD 16-MAY-1991.
PF 02-NOV-1990; U06548.
PR 03-NOV-1989; US-431565.
PA (IMMU-) IMMULOGIC PHARM COR.
PI Geffer ML, Garman RD, Greenstein JL, Juo M, Rogers BL;
PI Brauer AW;
PI

DR N-PSDB; Q11836.
PT New pure covalently linked human T cell reactive feline protein -
PT and modified peptide(s), used to reduce effects of cat allergens
PT and to diagnose sensitivity to allergens.
PS Claim 2; Fig 1; 70pp; English.
CC Poly-A mRNA from cat parotid and mandibular glands was used to
CC produce cDNA clones for both chain 1 and chain 2 of TRFP. These
CC clones were then used to screen a cat genomic library. Chain 1
CC exists in two forms having different leader sequences (A and B).
CC The sequence can be used to express the protein and peptide derivs.
CC which stimulate T-cells in persons allergic to cats. The peptides
CC can be used to reduce/eliminate the allergic response partic. by
CC modificn. of lymphokine prodn. by the T-cells. They can also be
CC used to identify epitopes responsible for sensitivity. The DNA can
CC be used to detect comparable sequence in other species, and also
CC for prodn. of modified forms of TRFP esp. showing reduced binding
CC to IgE and thus reduced tendency to cause adverse reactions.
CC See also R12120-R12123.
SQ Sequence 94 AA;

DB 3; Score 184; Match 100.0%; Predicted No. 4.68e-12;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 53 kalpvvlenarilkncvdakmteedke 79
|||||
Qy 1 KALPVVLENARILKNCVDAKMTEEDKE 27

RESULT 6

ID R36539 standard; Protein; 92 AA.
AC R36539;
DT 12-AUG-1993 (first entry)
DE TRFP chain 1 (with Leader A).
KW Human T cell reactive feline protein; TRFP; leader A; leader B;
KW epitope.
OS Felis.
FH Key Location/Qualifiers
FT Peptide 1..22
FT /label= leader_peptide
PN WD9308280-A.
PD 29-APR-1993.
PF 16-OCT-1992; U08694.
PR 16-OCT-1991; US-777859.
PR 13-DEC-1991; US-807529.
PA (IMMU-) IMMULOGIC PHARM CORP.
PI Bond JF, Garman RD, Kuo M, Morgenstern JP, Morville M;
PI Rogers BL;
DR WPI; 93-152473/18.
DR N-PSDB; Q41556.
PT Recombitope peptide having T-cell stimulating activity - for the
PT diagnosis and treatment of sensitivity to protein allergens,
PT auto:antigens and protein antigens
PS Disclosure; Fig 1; 73pp; English.
CC Chains 1 and 2 of the TRFP have been recombinantly expressed in E.
CC coli and purified. T cell epitope studies using overlapping peptide
CC regions derived from the TRFP amino acids sequence were used to
CC identify multiple T cell epitopes in each chain of TRFP.
SQ Sequence 92 AA;

DB 7; Score 184; Match 100.0%; Predicted No. 4.68e-12;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 51 kalpvvlenarilkncvdakmteedke 77
|||||
Qy 1 KALPVVLENARILKNCVDAKMTEEDKE 27

RESULT 7

ID R36543 standard; Protein; 27 AA.
 AC R36543;
 DT 12-AUG-1993 (first entry)
 DE Peptide Y.
 KW Human T cell reactive feline protein; TRFP; epitope; recombiteope.
 OS Felis.
 PN W09308280-A.
 PD 29-APR-1993.
 PF 16-OCT-1992; U08694.
 PR 16-OCT-1991; US-777859.
 PR 13-DEC-1991; US-807529.
 PA (IMMU-) IMMULOGIC PHARM CORP.
 PI Bond JF, Garman RD, Kuo M, Morgenstern JP, Morville M;
 PI Rogers BL;
 DR WPI; 93-152473/18.
 PT Recombiteope peptide having T-cell stimulating activity - for the
 PT diagnosis and treatment of sensitivity to protein allergens,
 PT auto:antigens and protein antigens
 PS Disclosure; Fig 4; 73pp; English.
 CC Chains 1 and 2 of the TRFP have been recombinantly expressed in E.
 CC coli and purified. T cell epitope studies using overlapping peptide
 CC regions derived from the TRFP amino acids sequence were used to
 CC identify multiple T cell epitopes in each chain of TRFP. DNA
 CC constructs were assembled in which 3 regions (encoding peptides X,
 CC Y and Z) were linked to produce DNA constructs encoding recombiteope-
 CC peptides.
 SQ Sequence 27 AA;

DB 7; Score 184; Match 100.0%; Predicted No. 4.68e-12;
 Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 kalpvvlenarilkncvdakmteedke 27
 |||||
 Qy 1 KALPVVLENARILKNCVDAKMTEEDKE 27

RESULT 8

ID R36540 standard; Protein; 88 AA.
 AC R36540;
 DT 12-AUG-1993 (first entry)
 DE TRFP chain 1 (with Leader B).
 KW Human T cell reactive feline protein; TRFP; leader A; leader B;
 KW epitope.
 OS Felis.
 FH Key Location/Qualifiers
 FT Peptide 1..18
 FT /label= leader_peptide
 PN W09308280-A.
 PD 29-APR-1993.
 PF 16-OCT-1992; U08694.
 PR 16-OCT-1991; US-777859.
 PR 13-DEC-1991; US-807529.
 PA (IMMU-) IMMULOGIC PHARM CORP.
 PI Bond JF, Garman RD, Kuo M, Morgenstern JP, Morville M;
 PI Rogers BL;
 DR WPI; 93-152473/18.
 DR N-PSDB; Q41557.
 PT Recombiteope peptide having T-cell stimulating activity - for the
 PT diagnosis and treatment of sensitivity to protein allergens,
 PT auto:antigens and protein antigens
 PS Disclosure; Fig 1; 73pp; English.
 CC Chains 1 and 2 of the TRFP have been recombinantly expressed in E.
 CC coli and purified. T cell epitope studies using overlapping peptide
 CC regions derived from the TRFP amino acids sequence were used to

CC identify multiple T cell epitopes in each chain of TRFP.
SQ Sequence 88 AA;

DB 7; Score 184; Match 100.0%; Predicted No. 4.68e-12;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 47 kalpvvlenarilkncvdakmteedke 73
|||||
Qy 1 KALPVVLENARILKNCVDAKMTEEDKE 27

RESULT 9

ID R12120 standard; Protein; 96 AA.
AC R12120;
DT 26-JUL-1991 (first entry)
DE TRFP chain 1 with leader B.
KW Human T cell reactive feline protein; cat allergens.
OS Felis catus.
FH Key Location/Qualifiers
FT Peptide 9..26
FT /label= Leader B
FT Protein 27..96
FT /label= TRFP Chain 1
PN W09106571-A.
PD 16-MAY-1991.
PF 02-NOV-1990; U06548.
PR 03-NOV-1989; US-431565.
PA (IMMU-) IMMULOGIC PHARM COR.
PI Gefter ML, Garman RD, Greenstein JL, Juo M, Rogers BL;
PI Brauer AW;
DR WPI; 91-164136/22.
DR N-PSDB; Q11837.
PT New pure covalently linked human T cell reactive feline protein -
PT and modified peptide(s), used to reduce effects of cat allergens
PT and to diagnose sensitivity to allergens.
PS Claim 2; Fig 1; 70pp; English.
CC Poly-A mRNA from cat parotid and mandibular glands was used to
CC produce cDNA clones for both chain 1 and chain 2 of TRFP. These
CC clones were then used to screen a cat genomic library. Chain 1
CC exists in two forms having different leader sequences (A and B).
CC The sequence can be used to express the protein and peptide derivs.
CC which stimulate T-cells in persons allergic to cats. The peptides
CC can be used to reduce/eliminate the allergic response partic. by
CC modifcn. of lymphokine prodn. by the T-cells. They can also be
CC used to identify epitopes responsible for sensitivity. The DNA can
CC be used to detect comparable sequence in other species, and also
CC for prodn. of modified forms of TRFP esp. showing reduced binding
CC to IgE and thus reduced tendency to cause adverse reactions.
CC See also R12119-R12123.
SQ Sequence 96 AA;

DB 3; Score 184; Match 100.0%; Predicted No. 4.68e-12;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 55 kalpvvlenarilkncvdakmteedke 81
|||||
Qy 1 KALPVVLENARILKNCVDAKMTEEDKE 27

RESULT 10

ID R27368 standard; protein; 96 AA.
AC R27368;
DT 25-FEB-1993 (first entry)
DE TRFP Chain #1 with C1 leader B sequence.
KW T cell reactive feline protein; cat allergy; allergic; IgE;
KW data release

OS Felis domesticus.
 FH Key Location/Qualifiers
 FT Peptide 1..27
 FT /label= Leader B
 FT Protein 28..96
 FT /label= TRFP chain #1
 PN W09215613-A.
 PD 17-SEP-1992.
 PF 20-FEB-1992; U01344.
 PR 28-FEB-1991; US-662193.
 PA (IMMU-) IMMULOGIC PHARM CORP.
 PI Bond J, Kuo M;
 DR WPI; 92-331670/40.
 PT Modified human T-cell reactive feline protein - stimulates T-cell
 PT in individuals allergic to cats and shows reduced
 PT histamine-releasing properties
 PS Claim 1; Fig 1; 35pp; English.
 CC This sequence represents a modified human T-cell reactive feline
 CC protein which stimulates T-cells from an individual who is allergic
 CC to cats, but which interacts with human IgE to a lesser extent than
 CC does affinity purified TRFP. The protein is modified by treating
 CC with either a mild alkali (pH 12.5-13.5 , KOH, NaOH, LiOH or tertiary
 CC amines) or an enzyme which removes O-linked groups (carbohydrate
 CC moieties). It is useful in desensitising people who are allergic to cats.
 SQ Sequence 96 AA;

DB 5; Score 184; Match 100.0%; Predicted No. 4.68e-12;
 Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 55 kalpvvlenarilkncvdakmteedke 81
 |||||
 Qy 1 KALPVVLENARILKNCVDAKMTEEDKE 27

RESULT 11

ID R27367 standard; protein; 94 AA.
 AC R27367;
 DT 25-FEB-1993 (first entry)
 DE TRFP Chain #1 with C1 leader A sequence.
 KW T cell reactive feline protein.
 OS Felis domesticus.
 FH Key Location/Qualifiers
 FT Peptide 1..25
 FT /label= Leader A
 FT Protein 25..94
 FT /label= TRFP chain #1
 PN W09215613-A.
 PD 17-SEP-1992.
 PF 20-FEB-1992; U01344.
 PR 28-FEB-1991; US-662193.
 PA (IMMU-) IMMULOGIC PHARM CORP.
 PI Bond J, Kuo M;
 DR WPI; 92-331670/40.
 PT Modified human T-cell reactive feline protein - stimulates T-cell
 PT in individuals allergic to cats and shows reduced
 PT histamine-releasing properties
 PS Claim 1; Fig 1; 35pp; English.
 CC This sequence represents a modified human T-cell reactive feline
 CC protein which stimulates T-cells from an individual who is allergic
 CC to cats, but which interacts with human IgE to a lesser extent than
 CC does affinity purified TRFP. The protein is modified by treating
 CC with either a mild alkali (pH 12.5-13.5 , KOH, NaOH, LiOH or tertiary
 CC amines) or an enzyme which removes O-linked groups (carbohydrate
 CC moieties). It is useful in desensitising people who are allergic to cats.
 SQ Sequence 94 AA;

DB 5; Score 184; Match 100.0%; Predicted No. 4.68e-12;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 53 kalpvvlenarilkncvdakmteedke 79
|||||
Qy 1 KALPVVLENARILKNCVDAKMTEEDKE 27

RESULT 12

ID P91948 standard; protein; 333 AA.
AC P91948;
DT 16-FEB-1990 (first entry)
DE Pig H4 isoenzyme.
KW NAD-dependent lactate dehydrogenase; H4 isoenzyme.
OS Suidae.
FH Key Location/Qualifiers
FT Binding-site 98..110
FT /note="substrate recognition site."
FT Binding-site 167..173
FT /note="activator site."
FT Misc-difference 102..102
FT /note="basic AA."
FT Misc-difference 173..173
FT /note="basic AA."
PN W08908707-A.
PD 21-SEP-1989.
PF 16-MAR-1989; G00279.
PR 17-MAR-1988; GB-006358.
PA (UYBR-) University of Bristol.
PI Holbrook JJ, Clarke AR, Atkinson A;
DR WPI; 89-292522/40.
PT Recombinant NAD-dependent dehydrogenase - which interconverts malate
PT and oxaloacetate, and has low dependence on fructose-1,6-biphosphate
PT as activator.
PS Disclosure; page 4-5; 25pp; English.
CC Sequence codes for the H4 isoenzyme of pig - an NAD-dependent lactate
CC dehydrogenase. It is used to construct a recombinant enzyme in which
CC AA102 and AA173 are basic, esp. Arg, and Gln resp. The mutation of AA102
CC results in the creation of a malate dehydrogenase from the lactate
CC dehydrogenase framework, the mutation being on the mobile coenzyme loop
CC and changing the substrate binding specificity of the protein. The
CC mutation of AA173, which is in the activation site, decreases
CC sensitivity of the protein to activation by sugar phosphates.
SQ Sequence 333 AA;

DB 1; Score 68; Match 37.5%; Predicted No. 6.74e+00;
Matches 9; Conservative 8; Mismatches 6; Indels 1; Gaps 1;

Db 290 slpcvl-nargltsvinqklkdde 312
: || || || | : : :
Qy 2 ALPVVLENARILKNCVDAKMTEED 25

RESULT 13

ID R43253 standard; Protein; 1513 AA.
AC R43253;
DT 04-MAY-1994 (first entry)
DE p190 protein.
KW p190; phosphoprotein; GTPase activating protein; GAP; n-chimerin;
KW mitogenically-stimulated cells; tyrosine kinase-transformed cells;
KW GAP-associated protein; homology; superfamily; signal transduction;
KW transcription repressor; GRF-1; BCR; breakpoint cluster region gene;
KW p21ras; effector; GTPase; mitogenic.
OS Rattus rattus.
FH Key Location/Qualifiers
FT

FT /note= "Region of homology to GTPase superfamily"
 FT Region 389..1166
 FT /note= "Region of homology to GRF-1"
 FT Region 1268..1429
 FT /note= "Region of homology to BCR/n-chimerin"
 PN W09J320201-A.
 PD 14-OCT-1993.
 PF 31-MAR-1993; U03076.
 PR 31-MAR-1992; US-861207.
 PA (WHED) WHITEHEAD INST BIOMEDICAL RES.
 PI Settleman JE; Weinberg RA;
 DR WPI; 93-336909/42.
 DR N-PSDB; Q50168.
 PT New GAP-associated protein P190 - which can be inhibited to
 PT interfere with RAS oncogene(s) in pathogenesis of malignancies
 PS Claim 4; Page 42-55; 95pp; English.
 CC This sequence represents the p190 protein. p190 is a phosphoprotein
 CC which is tightly bound to GTPase activating protein (GAP) in
 CC mitogenically-stimulated and tyrosine kinase-transformed cells.
 CC p190 is GAP-associated protein. p190 has three distinct domains,
 CC each of which exhibits homology to a previously described sequence.
 CC Towards the amino terminal end of p190, a domain spanning 201 amino
 CC acids exhibits significant sequence similarity to all members of the
 CC GTPase superfamily. The most striking similarity in the predicted
 CC amino acid sequence of p190, is to a 95 kD protein encoded by a human
 CC cDNA which is reported to function as a transcription repressor. The
 CC reported amino acid sequence of the repressor protein, GRF-1, is 95%
 CC identical over a 778 amino acid fragment of p190. Towards the carboxy
 CC terminal of p190 there is a region of 161 residues which shows homology
 CC with two proteins which are involved in signal transduction. It has
 CC been suggested that p21ras acts as a regulatory subunit of p190
 CC protein, which acts as the p21ras effector and which releases mitogenic
 CC signals when prompted to do so by activated GTP-bound p21ras. p190 may
 CC also, acting via GAP, transduce signals from p21ras to the nucleus,
 CC affecting expression of specific cellular genes.
 SQ Sequence 1513 AA;

DB 8; Score 60; Match 50.0%; Predicted No. 3.60e+01;
 Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Db 78 evrsrledcveckm 91
 | : | | : | | : | |
 Qy 8 ENARILKNCVDAKM 21

RESULT 14

ID R04571 standard; protein; 520 AA.
 AC R04571;
 DT 14-SEP-1990 (first entry)
 DE DRF3 product from the mos gene.
 KW Rhizopine; mos gene; moc gene; nitrogen fixation; Medicago sativa.
 OS Rhizobium meliloti strain L5-30.
 PN AU8941262-A.
 PD 15-MAR-1990.
 PF 08-SEP-1988; A41262.
 PR 08-SEP-1988; AU-000328.
 PA (LUMI-) Luminis PTY Ltd.
 PI Temp J, Kondorosi A, Putnoky P, Murphy PJ, Schell JS, De Bruijn FJ.
 DR WPI; 90-139827/19.
 DR N-PSDB; Q04303.
 PT Bacteria contg. genes for rhizopine synthesis and catabolism - esp.
 PT Rhizobium strains for increasing nitrogen fixation and growth in
 PT leguminous plants.
 PS Disclosure; p; English.
 CC The mos DRF 3 product is a protein of a predicted size of 35.8kD.
 CC Rhizobium strain of RM1021 contained the full mos gene and a mos gene

CC to catabolise rhizopine compounds are used to increase symbiotic nitrogen
CC fixation in Leguminaceae, esp. alfalfa. Where a *noc* gene is
CC present in separate bacteria both N-fixation and plant growth can be
CC promoted. Alternatively, *mos* genes are expressed in the plant and only
CC *noc* in the bacteria, this will cause desirable soil bacteria (eg being
CC used for biological control of a pathogen) to be held in the rhizosphere.
CC See also R04569-72.
SQ Sequence 520 AA;

DB 1: Score 60; Match 58.3%; Predicted No. 3.60e+01;
Matches 7; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

Db 4 lplvlqngqimk 15
||:|:|:|:|:
Qy 3 LPVVLENARILK 14

RESULT 15

ID P92275 standard; peptide; 765 AA.
AC P92275;
DT 27-Feb-1990 (first entry)
DE Human topoisomerase I cDNA
KW Scleroderma.
OS Homo sapiens (human).
PN W08909222-A.
PD 05-OCT-1989.
PF 22-MAR-1989; U01116.
PR 23-MAR-1988; US-172159.
PA (BRIG) Brigham and Women's Hospital; (UYJD) John's Hopkins Univ.
PI Earnshaw WC, D'Arpa P;
DR WPI; 89-309500/42.
DR N-PSDB; N91475.
PT Cloned cDNA encoding eukaryotic topoisomerase I - useful for large scale
PT prodn. by recombinant methods
PS Claim 6; fig. 5; 28pp; English.
CC The cDNA of this can be spliced into DNA vectors and used to transform
CC hosts for high yield. This polypeptide (I) retains the ability to bind
CC autoantibodies, even though the prokaryotic host degrades transcribed (I)
CC into a spectrum of polypeptides. (I) may be used to classify patients
CC with immune rheumatic diseases.
SQ Sequence 765 AA;

DB 1: Score 59; Match 50.0%; Predicted No. 4.41e+01;
Matches 10; Conservative 5; Mismatches 4; Indels 1; Gaps 1;

Db 445 etarrlkkcvd-kirngyre 463
| || ||:| || | : : : |
Qy 8 ENARILKNCVDAKMTEEDKE 27

Search completed: Fri Mar 24 07:44:12 1995
Job time : 11 secs.

MIRSLA (TM)

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Fri Mar 24 07:43:20 1995; MasPar time 4.72 Seconds
128.633 Million cell updates/sec

Tabular output not generated.

Title: >US-08-300-510-2
Description: (1:27) from US08300510.pep
Perfect Score: 184
Sequence: 1 KALPVVLENARILKNCVDAKMTEEDKE 27

Scoring table: PAM 150
Gap 14

Searched: 75511 seqs, 22468834 residues

Database: pir43
1 ANNO1
2 ANNO2
3 ANNO3
4 UNANNO1
5 UNANNO2
6 UNANNO3
7 UNANNO4
8 UNANNO5
9 UNANNO6
10 UNREV1
11 UNREV2
12 UNREV3

Statistics: Mean 29.063; Variance 55.105; scale 0.527

Predicted No. is the number of results expected by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	178	96.7	88	9	JC1126	major allergen chain	1.31e-19
2	178	96.7	92	9	JC1136	major allergen chain	1.31e-19
3	80	43.5	40	9	A53283	major cat allergen F	4.90e-02
4	79	42.9	2333	3	GNNY2F	genome polyprotein -	6.98e-02
5	75	40.8	2336	5	S37077	genome polyprotein -	2.80e-01
6	72	39.1	2332	3	GNNYF	genome polyprotein -	7.76e-01
7	71	38.6	136	7	S38598	hypothetical protein	1.09e+00
8	71	38.6	374	8	S28285	hypothetical protein	1.09e+00
9	70	38.0	386	11	S37691	rnalp protein - fiss	1.51e+00
10	69	37.5	470	5	S02068	RNA-directed RNA pol	2.10e+00
11	69	37.5	470	5	JN0431	RNA-directed RNA pol	2.10e+00
12	68	37.0	333	1	DEPGLH	L-lactate dehydrogen	2.91e+00
13	66	35.9	217	4	B32957	L-lactate dehydrogen	5.55e+00
14	66	35.9	334	4	S02795	L-lactate dehydrogen	5.55e+00
15	66	35.9	334	4	S09954	L-lactate dehydrogen	5.55e+00
16	66	35.9	2332	3	GNNY4F	genome polyprotein -	5.55e+00
17	66	35.9	1830	2	S19188	myosin-V - chicken	5.55e+00
18	64	34.8	147	5	S07158	keratin type I compo	1.05e+01
19	64	34.8	412	2	KRSHL1	keratin, 48K type I	1.05e+01
20	64	34.8	314	10	S31402	3-methylcatechol 2,3	1.05e+01
21	63	34.2	529	5	A24031	genome polyprotein -	1.43e+01
22	63	34.2	4427	10	S25021	probable polyketide	1.43e+01
23	-	-	90	7	UA074	acetyl-CoA hydrolase	1.43e+01

24	63	34.2	906	7	S32607	trifunctional enzyme	1.43e+01
25	63	34.2	906	7	JT0350	hydratase (EC 4.2.1.	1.43e+01
26	62	33.7	419	5	A25438	keratin, type I cyto	1.95e+01
27	62	33.7	404	5	JS0073	keratin, 47.6K type	1.95e+01
28	61	33.2	873	1	TVFVFS	protein-tyrosine kin	2.64e+01
29	61	33.2	873	1	TVFVF	protein-tyrosine kin	2.64e+01
30	61	33.2	90	6	A49923	translation initiati	2.64e+01
31	61	33.2	277	6	JN0751	Outer membrane 30K p	2.64e+01
32	60	32.6	1226	11	S48837	kinesin-like protein	3.57e+01
33	60	32.6	1493	9	A38218	GAP-associated prote	3.57e+01
34	59	32.1	900	7	S25322	bifunctional beta-ox	4.81e+01
35	59	32.1	767	12	S32698	DNA topoisomerase (E	4.81e+01
36	59	32.1	767	12	S32697	DNA topoisomerase (E	4.81e+01
37	59	32.1	135	11	S46635	hypothetical protein	4.81e+01
38	59	32.1	765	1	ISHUT1	DNA topoisomerase (E	4.81e+01
39	59	32.1	767	4	JU0144	DNA topoisomerase (E	4.81e+01
40	59	32.1	567	6	A40899	gag polyprotein - Ch	4.81e+01
41	59	32.1	400	8	A61556	keratin 19, cytoskel	4.81e+01
42	59	32.1	631	9	A31203	interferon-regulated	4.81e+01
43	59	32.1	416	9	A61404	keratin A, type I -	4.81e+01
44	59	32.1	1178	9	A47255	pyruvate carboxylase	4.81e+01
45	59	32.1	698	3	IKEC5B	colicin V secretion	4.81e+01

ALIGNMENTS

```

RESULT      1
ENTRY       JC1126      #type complete
TITLE       major allergen chain 1 precursor B - cat
ORGANISM    #formal_name Felis silvestris catus #common_name domestic cat
DATE        31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change
            31-Dec-1993
ACCESSIONS  JC1126
REFERENCE   JC1126
#authors    Griffith, I.J.; Craig, S.; Pollock, J.; Yu, X.B.;
            Morgenstern, J.P.; Rogers, B.L.
#journal     Gene (1992) 113:263-268
#title       Expression and genomic structure of the genes encoding FdI,
            the major allergen from the domestic cat.
#accession   JC1126
##molecule_type DNA
##residues   1-88 ##label GRI
GENETICS
#gene        Ch1
#introns     17/1; 79/3
FEATURE
1-18         #domain signal sequence #status predicted #label SIG\
19-88        #product major allergen chain 1 #status predicted #label
            MAT
SUMMARY      #length 88 #molecular-weight 9586 #checksum 4095

DB 9; Score   178; Match 96.3%; Predicted No. 1.31e-19;
Matches 26; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 47 nalpvvlenarilkncvdakmteedke 73
:|||||
Qy 1 KALPVVLENARILKNCVDAKMTEEDKE 27

```

```

RESULT      2
ENTRY       JC1136      #type complete
TITLE       major allergen chain 1 precursor A - cat
ORGANISM    #formal_name Felis silvestris catus #common_name domestic cat
DATE        31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change
            31-Dec-1993

```

REFERENCE JC1126
 #authors Griffith, I.J.; Craig, S.; Pollock, J.; Yu, X.B.;
 Morgenstern, J.P.; Rogers, B.L.
 #journal Gene (1992) 113:263-268
 #title Expression and genomic structure of the genes encoding FdI,
 the major allergen from the domestic cat.
 #accession JC1136
 ##molecule_type DNA
 ##residues 1-92 ##label GRI

GENETICS
 #gene Ch1
 #introns 21/1; 83/3

FEATURE
 1-22 #domain signal sequence #status predicted #label SIG\
 23-92 #product major allergen chain 1 #status predicted #label
 MAT

SUMMARY #length 92 #molecular-weight 10072 #checksum 4988

DB 9; Score 178; Match 96.3%; Predicted No. 1.31e-19;
 Matches 26; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 51 nalpvvlenarilkncvdakmteedke 77
 :|
 @y 1 KALPVVLENARILKNCVDAKMTEEDKE 27

RESULT 3

ENTRY A53283 #type fragment
 TITLE major cat allergen Fel d 1 alpha chain - cat (fragment)
 ORGANISM #formal_name Felis silvestris catus #common_name domestic cat
 DATE 12-May-1994 #sequence_revision 12-May-1994 #text_change
 12-May-1994

ACCESSIONS A53283
 REFERENCE A53283
 #authors Duffort, D.A.; Carreira, J.; Nitti, G.; Polo, F.; Lombardero,
 M.
 #journal Mol. Immunol. (1991) 28:301-309
 #title Studies on the biochemical structure of the major cat
 allergen Felis domesticus I.
 #accession A53283
 ##status preliminary
 ##molecule_type protein
 ##residues 1-40 ##label DUF

SUMMARY #length 40 #checksum 3032

DB 9; Score 80; Match 100.0%; Predicted No. 4.90e-02;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 29 kalpvvlenari 40
 ||
 @y 1 KALPVVLENARI 12

RESULT 4

ENTRY GNNY2F #type complete
 TITLE genome polyprotein - foot-and-mouth disease virus A (strain
 AC10161)
 CONTAINS coat protein VP1; coat protein VP2; coat protein VP3; coat
 protein VP4; core protein p52; genome-linked protein VPg1;
 genome-linked protein VPg2; genome-linked protein VPg3;
 nonstructural protein p20a; nonstructural protein p20b;
 RNA-directed RNA polymerase (EC 2.7.7.48)
 ORGANISM #formal_name Aphthovirus A #common_name foot-and-mouth
 disease virus A
 DATE 17-Dec-1982 #sequence_revision 28-Aug-1985 #text_change
 21-Dec-1997

ACCESSIONS A93508; A91491; S30753
 REFERENCE A93508
 #authors Carroll, A.R.; Rowlands, D.J.; Clarke, B.E.
 #journal Nucleic Acids Res. (1984) 12:2461-2472
 #title The complete nucleotide sequence of the RNA coding for the
 primary translation product of foot and mouth disease
 virus.
 #cross-references MUID:84169547
 #accession A93508
 ##molecule_type genomic RNA
 ##residues 1-2333 ##label CAR
 ##cross-references GB:X00429
 REFERENCE A91491
 #authors Boothroyd, J.C.; Harris, T.J.R.; Rowlands, D.J.; Lowe, P.A.
 #journal Gene (1982) 17:153-161
 #title The nucleotide sequence of cDNA coding for the structural
 proteins of foot-and-mouth disease virus.
 #cross-references MUID:82211814
 #accession A91491
 ##molecule_type genomic RNA
 ##residues 115-395,'C',397-631,'L',633-1048 ##label B00
 ##cross-references GB:V01130
 REFERENCE S30753
 #authors Sangar, D.V.; Newton, S.E.; Rowlands, D.J.; Clarke, B.E.
 #journal Nucleic Acids Res. (1987) 15:3305-3315
 #title All foot and mouth disease virus serotypes initiate protein
 synthesis at two separate AUGs.
 #accession S30753
 ##molecule_type genomic RNA
 ##residues 1-32 ##label SAN
 ##cross-references EMBL:M31575
 CLASSIFICATION #superfamily foot-and-mouth disease virus genome polyprotein
 KEYWORDS coat protein; core protein; genome-linked protein;
 nonstructural protein; nucleotidyltransferase; polyprotein
 FEATURE
 1-204 #product nonstructural protein p20a #label NPA\
 205-286 #product coat protein VP4 #label VP4\
 287-504 #product coat protein VP2 #label VP2\
 505-725 #product coat protein VP3 #label VP3\
 726-937 #product coat protein VP1 #label VP1\
 938-1578 #product core protein p52 #label CPP\
 1579-1601 #product genome-linked protein VPg1 #label GL1\
 1602-1625 #product genome-linked protein VPg2 #label GL2\
 1626-1649 #product genome-linked protein VPg3 #label GL3\
 1650-1863 #product nonstructural protein p20b #label NPB\
 1864-2333 #product RNA-directed RNA polymerase #label RRP
 SUMMARY #length 2333 #molecular-weight 259646 #checksum 7155
 DB 3; Score 79; Match 59.1%; Predicted No. 6.98e-02;
 Matches 13; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

Db 1913 vvlddvifskhkgdakmteedk 1934
 |||:: :|: |||||
 @y 5 VVLENARILKNCVDKMTTEEDK 26

RESULT 5
 ENTRY S37077 #type complete
 TITLE genome polyprotein - foot-and-mouth disease virus A (strain
 A22/550 Azerbaijan 65)
 CONTAINS coat protein VP1; coat protein VP2; coat protein VP3; coat
 protein VP4; core protein p14; core protein p19; core
 protein p41; core protein X; genome-linked protein VPg1;
 genome-linked protein VPg2; genome-linked protein VPg3;
 nonstructural protein p20a; proteinase (EC 3.4.-.-);
 RNA-directed RNA polymerase (EC 2.7.7.48)

ORGANISM #formal_name Aphthovirus A #common_name foot-and-mouth disease virus A
 DATE 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 31-Dec-1993
 ACCESSIONS 537077
 REFERENCE 537077
 #authors Sosnovtsev, S.V.; Onischenko, A.M.; Petrov, N.A.; Kalashnikova, T.I.; Mamaeva, N.V.; Drygin, V.Y.; Perevozchikova, N.A.; Vasilenko, S.K.
 #submission submitted to the EMBL Data Library, August 1993
 #accession 537077
 ##molecule_type genomic RNA
 ##residues 1-2336 ##label SOS
 ##cross-references EMBL:X74812
 CLASSIFICATION #superfamily foot-and-mouth disease virus genome polyprotein
 KEYWORDS coat protein; core protein; genome-linked protein; nonstructural protein; nucleotidyltransferase; polyprotein
 FEATURE
 1-217 #product nonstructural protein p20a #status predicted #label NPA\
 218-286 #product coat protein VP4 #status predicted #label VP4\
 287-504 #product coat protein VP2 #status predicted #label VP2\
 505-724 #product coat protein VP3 #status predicted #label VP3\
 725-938 #product coat protein VP1 #status predicted #label VP1\
 939-954 #product core protein X #status predicted #label CPX\
 955-1108 #product core protein p14 #status predicted #label C14\
 1109-1426 #product core protein p41 #status predicted #label C41\
 1427-1579 #product core protein p19 #status predicted #label C19\
 1580-1602 #product genome-linked protein VPg1 #status predicted #label VG1\
 1603-1626 #product genome-linked protein VPg2 #status predicted #label VG2\
 1627-1650 #product genome-linked protein VPg3 #status predicted #label VG3\
 1651-1863 #product proteinase #status predicted #label PTS\
 1864-2333 #product RNA-directed RNA polymerase #status predicted #label RRP
 SUMMARY #length 2336 #molecular-weight 259983 #checksum 4399
 DB 5; Score 75; Match 54.5%; Predicted No. 2.80e-01;
 Matches 12; Conservative 5; Mismatches 5; Indels 0; Gaps 0;
 Db 1913 vvldevifskhkgtkmtedk 1934
 |||:: : | : |||||
 Qy 5 VVLENARILKNCVDAKMTEEDK 26
 RESULT 6
 ENTRY GNNYF #type complete
 TITLE genome polyprotein - foot-and-mouth disease virus O (strains O1K and O1BFS)
 CONTAINS coat protein VP1; coat protein VP2; coat protein VP3; coat protein VP4; core protein p12; core protein p14; core protein P20b; core protein p34; core protein P56; core protein VPg; nonstructural protein p20a
 ORGANISM #formal_name Aphthovirus O #common_name foot-and-mouth disease virus O
 #note host Artiodactyla (cloven-footed mammals)
 DATE 01-Sep-1981 #sequence_revision 27-Nov-1985 #text_change 08-Apr-1994
 ACCESSIONS A03907; A37503
 REFERENCE A03907
 #authors Forss, S.; Strebel, K.; Beck, E.; Schaller, H.
 #journal Nucleic Acids Res. (1984) 12:6587-6601
 #title Nucleotide sequence and genome organization of foot-and-mouth disease virus

```

#cross-references MUID:84297249
#contents      strain 01K
#accession     A03907
  ##molecule_type mRNA
  ##residues    1-2332 ##label FOR
REFERENCE      A37503
  #authors      Makoff, A.J.; Paynter, C.A.; Rowlands, D.J.; Boothroyd, J.C.
  #journal      Nucleic Acids Res. (1982) 10:8285-8295
  #title        Comparison of the amino acid sequence of the major immunogen
                from three serotypes of foot and mouth disease virus.
#cross-references MUID:83143292
#contents      strain 01BFS
#accession     A37503
  ##molecule_type genomic RNA
  ##residues    715-779,'V',781-807,'R',809-860,'S',862-951 ##label MAK
COMMENT        The coat protein VP1 contains the main antigenic determinants of
                the virion; therefore, changes in its sequence must be
                responsible for the high antigenic variability of the virus.
COMMENT        Coat proteins VP2 and VP3 are related to the poliovirus coat
                proteins VP2 and VP3.
CLASSIFICATION #superfamily foot-and-mouth disease virus genome polyprotein
KEYWORDS       coat protein; core protein; nonstructural protein;
                polyprotein
FEATURE
  1-217         #product nonstructural protein p20a #label NPA\
  218-286       #product coat protein VP4 #label VP4\
  287-504       #product coat protein VP2 #label VP2\
  505-724       #product coat protein VP3 #label VP3\
  725-937       #product coat protein VP1 #label VP1\
  938-1107      #product core protein p12 #label C12\
  1108-1425     #product core protein p34 #label P34\
  1426-1578     #product core protein p14 #label C14\
  1579-1649     #product genome-linked protein VPg #label VPG\
  1650-1862     #product nonstructural protein p20b #label P20\
  1863-2332     #product RNA-directed RNA polymerase #label P56
SUMMARY        #length 2332 #molecular-weight 258925 #checksum 4170

DB 3; Score      72; Match 50.0%; Predicted No. 7.76e-01;
Matches 11; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

Db 1912 vvldevifskhkgtkmseedk 1933
   |||:: : | : |::|
Qy 5 VVLENARILKNCVDAKMTEEDK 26

RESULT 7
ENTRY   S38598 #type fragment
TITLE   hypothetical protein 136 (rpl20 5' region) - euglenid
        (Astasia longa) plastid (fragment)
ORGANISM #formal_name plastid Astasia longa
DATE     31-Dec-1993 #sequence_revision 02-Aug-1994 #text_change
        02-Aug-1994
ACCESSIONS S38598
REFERENCE  S38590
  #authors  Gockel, G.; Baier, S.; Hachtel, W.
  #submission submitted to the EMBL Data Library, November 1993
  #accession S38598
  ##molecule_type DNA
  ##residues 1-136 ##label GOC
  ##cross-references EMBL:X75653
KEYWORDS  plastid
SUMMARY   #length 136 #checksum 6797

DB 7; Score      71; Match 46.7%; Predicted No. 1.09e+00;
Matches 7; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

```

Q4. 7 LENARILKNCVDAKM 21

RESULT 8

```
ENTRY      S28285      #type complete
TITLE      hypothetical protein C38C10.1 - Caenorhabditis elegans
ORGANISM    #formal_name Caenorhabditis elegans
DATE        12-Mar-1993 #sequence_revision 12-Mar-1993 #text_change
            30-Sep-1993
```

```

ACCESSIONS      528285
REFERENCE       528285
#authors        Thomas, K.
#submission     submitted to the EMBL Data Library, December 1992
#accession      528285
##molecule_type DNA
##residues      1-374 ##label THO
##cross-references EMBL:Z19153

```

GENETICS
#introns 8/2; 108/2; 149/3; 176/2; 225/3; 289/2; 349/1
SUMMARY #length 374 #molecular-weight 42940 #checksum 2438

DB 8; Score 71; Match 22.2%; Predicted No. 1.09e+00;
Matches 6; Conservative 14; Mismatches 7; Indels 0; Gaps 0;

```
Db 322 rsmalslqkgrvnsscldkkvkenssq 348
      :::: |::| : |::| |::| :
Qu 1 KALPVLLENARILKNCVDAKMTEEDKE 27
```

RESULT 9

```
ENTRY      537691      #type complete
TITLE      rnalp protein - fission yeast (Schizosaccharomyces pombe)
ORGANISM    #formal_name Schizosaccharomyces pombe
DATE        18-May-1994; #sequence_revision 18-May-1994; #text_change
            18-May-1994
```

```

ACCESSIONS      S37691
REFERENCE       S37691
#authors        Melchior, F.; Weber, K.; Gerke, V.
#journal        Mol. Biol. Cell (1993) 4:569-581
#title          A functional homologue of the RNA1 gene product in
                  Schizosaccharomyces pombe; purification, biochemical
                  characterization, and identification of a leucine-rich
                  repeat motif.
#accession      S37691
##status        preliminary
##residues      1-386 ##label MEL
##cross-references EMBL:X69882
SUMMARY         #length 386 #molecular-weight 43235 #checksum 8326

```

DB 11; Score 70; Match 36.8%; Predicted No. 1.51e+00;
Matches 7; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

Db 286 ieldavrtlktvidekmpd 304
: |: | | :| | :
Qu 5 VVLENARILKNCVDKMT 23

RESULT 10

```

ENTRY      S02068      #type complete
TITLE      RNA-directed RNA polymerase (EC 2.7.7.48) - foot-and-mouth
            disease virus A
ALTERNATE_NAMES RNA replicase
ORGANISM     #formal_name Aphthovirus A #common_name foot-and-mouth
            disease virus A

```

DATE 01-Dec-1989 #sequence_revision 01-Dec-1989 #text_change
30-Sep-1993

ACCESSIONS 502068

REFERENCE 502068

#authors Villaverde, A.; Martinez-Salas, E.; Domingo, E.

#journal J. Mol. Biol. (1988) 204:771-776

#title 3D gene of foot-and-mouth disease virus. Conservation by
convergence of average sequences.

#cross-references MUID:89141768

#accession 502068

##molecule_type mRNA

##residues 1-470 ##label VIL

##note 48-Gly, 68-Ala, 158-Val, 274-Ile, 306-Ile, 374-Leu, and
444-Glu were also found

##note sequence not compared to nucleotide translation

GENETICS

#gene 3D

CLASSIFICATION #superfamily foot-and-mouth disease virus genome polyprotein

KEYWORDS nucleotidyltransferase

SUMMARY #length 470 #molecular-weight 52910 #checksum 502

DB 5; Score 69; Match 45.5%; Predicted No. 2.10e+00;
Matches 10; Conservative 7; Mismatches 5; Indels 0; Gaps 0;

Db 50 vvldevifsrhkgdtkmseedk 71
|||: : : |:|:|
Qy 5 VVLENARILKNCVDAKMTEEDK 26

RESULT 11

ENTRY JN0431 #type complete

TITLE RNA-directed RNA polymerase (EC 2.7.7.48) - foot-and-mouth
disease virus A (strain A22)

ORGANISM #formal_name Aphthovirus A #common_name foot-and-mouth
disease virus A

DATE 05-Mar-1993 #sequence_revision 05-Mar-1993 #text_change
30-Sep-1993

ACCESSIONS JN0431

REFERENCE JN0431

#authors Kuzmin, I.V.; Rybakov, S.S.; Ivanyushchenkov, V.N.; Burdov,
A.N.

#journal Bioorg. Khim. (1989) 15:419-422

#title Nucleotide sequence of the FMDV A22 RNA polymerase gene.

#cross-references MUID:89302183

#accession JN0431

##molecule_type mRNA

##residues 1-470 ##label KUZ

##note this paper is in Russian, with an English abstract

CLASSIFICATION #superfamily foot-and-mouth disease virus genome polyprotein

KEYWORDS nucleotidyltransferase

SUMMARY #length 470 #molecular-weight 52657 #checksum 1182

DB 5; Score 69; Match 50.0%; Predicted No. 2.10e+00;
Matches 11; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

Db 50 vvldevifskhkgdtkmtaek 71
|||: : |:|:|
Qy 5 VVLENARILKNCVDAKMTEEDK 26

RESULT 12

ENTRY DEPLGH #type complete

TITLE L-lactate dehydrogenase (EC 1.1.1.27) chain H - pig

ORGANISM #formal_name Sus scrofa domestica #common_name domestic pig

DATE #sequence_revision 07-May-1981 #text_change 05-Aug-1994

ACCESSIONS


```

REFERENCE      A91671
#authors      Kiltz, H.H.; Keil, W.; Griesbach, M.; Petry, K.; Meyer, H.
#journal      Hoppe-Seyler's Z. Physiol. Chem. (1977) 358:123-127
#title        The primary structure of porcine lactate dehydrogenase:
               isoenzymes M-4 and H-4.
#cross-references MUID:77117453
#accession    A91671
               ##molecule_type protein
               ##residues      1-20,'B',22-146,'A',148-214,'B',216,'Z',218-333 ##label
               KIL

REFERENCE      A94603
#authors      Kiltz, H.H.
#submission    submitted to the Atlas, October 1977
#accession    A94603
               ##molecule_type protein
               ##residues      1-333 ##label KI2

REFERENCE      A92870
#authors      Grau, U.M.; Trommer, W.E.; Rossmann, M.G.
#journal      J. Mol. Biol. (1981) 151:289-307
#title        Structure of the active ternary complex of pig heart lactate
               dehydrogenase with S-lac-NAD at 2.7 angstrom resolution.
#cross-references MUID:82170431
#contents      annotation; X-ray crystallography, 2.7 angstroms
#note          the structure of a complex with a coenzyme-substrate analog
               was solved

COMMENT        A tetramer of H chains is the predominant form of the enzyme in
               heart muscle.

CLASSIFICATION #superfamily L-lactate dehydrogenase
KEYWORDS        acetylated amino end; NAD; oxidoreductase; tetramer
FEATURE
  1              #modified_site acetylated amino end (Ala) #status
               experimental\
  163            #active_site Cys #status experimental
SUMMARY          #length 333 #molecular-weight 36476 #checksum 6356

DB 1; Score      68; Match 37.5%; Predicted No. 2.91e+00;
Matches 9; Conservative 8; Mismatches 6; Indels 1; Gaps 1;

Db 290 slpcvl-nargltsvinqklkdde 312
   :|| || ||| | : :: | : ::
Qy 2 ALPVVLENARILKNCVDAKMTEED 25

RESULT 13
ENTRY      B32957 #type fragment
TITLE      L-lactate dehydrogenase (EC 1.1.1.27) chain H - rabbit
               (fragment)
ORGANISM    #formal_name Oryctolagus cuniculus #common_name domestic
               rabbit
DATE        22-Nov-1989 #sequence_revision 22-Nov-1989 #text_change
               02-Aug-1994
ACCESSIONS  B32957
REFERENCE   A32957
#authors    Sass, C.; Briand, M.; Benslimane, S.; Renaud, M.; Briand, Y.
#journal    J. Biol. Chem. (1989) 264:4076-4081
#title      Characterization of rabbit lactate dehydrogenase-M and
               lactate dehydrogenase-H cDNAs. Control of lactate
               dehydrogenase expression in rabbit muscle.
#cross-references MUID:89139477
#accession    B32957
               ##status      preliminary
               ##molecule_type mRNA
               ##residues      1-217 ##label SAS
               ##cross-references GB:M22584; GB:J04595
CLASSIFICATION #superfamily L-lactate dehydrogenase

```

SUMMARY

#length 217 #checksum 4425

DB 4; Score 66; Match 33.3%; Predicted No. 5.55e+00;
Matches 8; Conservative 9; Mismatches 6; Indels 1; Gaps 1;

Db 174 slpcil-nargltsvinqklkdde 196
: || : | ||| | : : : | : : :
Qy 2 ALPVVLENARILKNCVDAKMTEED 25

RESULT 14

ENTRY 502795 #type complete
TITLE L-lactate dehydrogenase (EC 1.1.1.27) B - human
ORGANISM #formal_name Homo sapiens #common_name man
DATE 01-Dec-1989 #sequence_revision 01-Dec-1989 #text_change
02-Aug-1994
ACCESSIONS 502795; S06281
REFERENCE 502795
#authors Takeno, T.; Li, S.S.L.
#journal Biochem. J. (1989) 257:921-924
#title Structure of the human lactate dehydrogenase B gene.
#cross-references MUID:89193506
#accession 502795
##molecule_type DNA
##residues 1-334 ##label TAK
##cross-references EMBL:X13794
REFERENCE S06281
#authors Sakai, I.; Sharief, F.S.; Pan, Y.C.E.; Li, S.S.L.
#journal Biochem. J. (1987) 248:933-936
#title The cDNA and protein sequences of human lactate dehydrogenase
B.
#cross-references MUID:88133965
#accession S06281
##molecule_type mRNA
##residues 1-334 ##label SAK
##note part of this sequence was confirmed by protein
sequencing

GENETICS

#gene GDB:LDHB
#map_position 12p12.2-p12.1
#introns 43/3; 83/1; 141/1; 199/1; 238/2; 279/3
CLASSIFICATION #superfamily L-lactate dehydrogenase
KEYWORDS NAD; oxidoreductase
FEATURE
2-334 #product L-lactate dehydrogenase B #status predicted
#label MAT
SUMMARY #length 334 #molecular-weight 36638 #checksum 6440

DB 4; Score 66; Match 33.3%; Predicted No. 5.55e+00;
Matches 8; Conservative 9; Mismatches 6; Indels 1; Gaps 1;

Db 291 slpcil-nargltsvinqklkdde 313
: || : | ||| | : : : | : : :
Qy 2 ALPVVLENARILKNCVDAKMTEED 25

RESULT 15

ENTRY 509954 #type complete
TITLE L-lactate dehydrogenase (EC 1.1.1.27) B - mouse
ORGANISM #formal_name Mus musculus #common_name house mouse
DATE 12-Feb-1993 #sequence_revision 12-Feb-1993 #text_change
02-Aug-1994
ACCESSIONS 509954
REFERENCE 509954
#authors Hiraoka, B.Y.; Sharief, F.S.; Yang, Y.W.; Li, W.H.; Li,
c_s

```
#journal      Eur. J. Biochem. (1990) 189:215-220
#title       The cDNA and protein sequences of mouse lactate dehydrogenase
              B. Molecular evolution of vertebrate lactate dehydrogenase
              genes A (muscle), B (heart) and C (testis).
```

#cross-references MUID:90249362

#accession 609954

```
##molecule_type mRNA
```

```
##residues      1-334  ##label HIR
```

##cross-references EMBL:X51905

```
##cross-references ENDE+X01700
##note      the authors translated the codon CTG for residue 41 as
```

Lys and AAT for residue 306 as Asp

CLASSIFICATION #superfamily L-lactate dehydrogenase

KEYWORDS NAD; oxidoreductase

KEYWORDS NAD; Oxidoreductase

SUMMARY #length 334 #molecular-weight 36572 #checksum 6533

DB 4; Score 66; Match 33.3%; Predicted No. 5.55e+00;

DB 4; Score 66; Match 33.3%; Predicted No. Sites 66;
Matches 8; Conservative 9; Mismatches 6; Indels 1; Gaps 1;

Db 291 slpcil-nargltsvinqklkdde 313

111 71 111 1 7 77 7 777

2 ALPVVLENARILKNCVDAKMTED 25

Search completed: Fri Mar 24 07:43:41 1995

Job time : 21 secs.

The logo consists of the word "NISELE" in a bold, blocky, sans-serif font. Each letter is formed by a thick dashed outline. The letters are closely spaced. To the right of the word is a small trademark symbol (TM).

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

Release 2.0 John F. Collins & S. S. Sturrock, Biocomputing Research Unit.

Copyright (c) 1993, 1994 by University of Edinburgh, U.K.

Distribution rights by IntelliGenetics, Inc.

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

```
Run on:      Fri Mar 24 07:42:48 1995;  MasPar time 3.42 Seconds
           111.824 Million cell updates/sec
```

Tabular output not generated.

Title: ~~US-08-300-510-2~~

Description: (1:27) from US08300510.ppt

Perfect Score: 184

Sequence: 1 KALPVLLENARILKNCVDAKMTEEDKE 27

Scoring table: PAM 150

Gap 14

Searched: 40292 seqs, 14147368 residues

Database: swiss-prot30

1 SPT 1

2 SPT2

3 SPT3

4 SPT4

5 SPT5

6 SFT6

7 SPT7

Statistics: Mean 30.782; Variance 44.012; scale 0.699

Predicted No. is the number of results expected by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description	Pred. No.
1	184	100.0	88	2	FELB_FELCA	MAJOR ALLERGEN I POLY	1.03e-27
2	184	100.0	92	2	FELA_FELCA	MAJOR ALLERGEN I POLY	1.03e-27
3	79	42.9	2333	5	POLG_FMDV1	GENOME POLYPROTEIN (N	3.09e-03
4	72	39.1	2332	5	POLG_FMDV0	GENOME POLYPROTEIN (N	6.49e-02
5	71	38.6	374	7	YLD1_CAEEL	PROBABLE G PROTEIN-CO	9.89e-02
6	71	38.6	136	7	YCTP_ASTLD	HYPOTHETICAL PROTEIN	9.89e-02
7	68	37.0	333	4	LDHH_PIG	L-LACTATE DEHYDROGENA	3.42e-01
8	66	35.9	333	4	LDHH_HUMAN	L-LACTATE DEHYDROGENA	7.66e-01
9	66	35.9	2332	5	POLG_FMDVA	GENOME POLYPROTEIN (N	7.66e-01
10	66	35.9	217	4	LDHH_RABIT	L-LACTATE DEHYDROGENA	7.66e-01
11	66	35.9	333	4	LDHH_MOUSE	L-LACTATE DEHYDROGENA	7.66e-01
12	66	35.9	1829	4	MYSD_CHICK	DILUTE MYOSIN HEAVY C	7.66e-01
13	64	34.8	215	1	ACRR_ECOLI	POTENTIAL REPRESSOR F	1.69e+00
14	63	34.2	906	3	HDE_CANTR	HYDRATASE-DEHYDROGENA	2.49e+00
15	63	34.2	861	5	POLG_FMDVS	GENOME POLYPROTEIN (C	2.49e+00
16	62	33.7	419	4	K1C4_XENLA	KERATIN, TYPE I CYTOS	3.65e+00
17	62	33.7	412	4	K1M1_SHEEP	KERATIN, TYPE I MICRO	3.65e+00
18	62	33.7	403	4	K1M2_SHEEP	KERATIN, TYPE I MICRO	3.65e+00
19	62	33.7	247	6	SUMT_PSEFL	UROPORPHYRIN-III C-ME	3.65e+00
20	61	33.2	90	3	IF1_CHLTR	INITIATION FACTOR IF-	5.33e+00
21	61	33.2	873	4	KFPS_FUJSV	TYROSINE-PROTEIN KINA	5.33e+00
22	59	32.1	400	4	K1CS_HUMAN	KERATIN, TYPE I CYTOS	1.12e+01
23	59	32.1	698	2	CVAB_ECOLI	COLICIN V SECRETION A	1.12e+01
24	59	32.1	767	6	TOP1_MOUSE	DNA TOPOISOMERASE I (1.12e+01
25	59	32.1	1178	5	PYC_MOUSE	PYRUVATE CARBOXYLASE	1.12e+01
26	59	32.1	1177	5	PH81_YEAST	PHOSPHATE SYSTEM POSI	1.12e+01
27	59	32.1	765	6	TOP1_HUMAN	DNA TOPOISOMERASE I (1.12e+01
28	59	32.1	631	4	MX1_MOUSE	INTERFERON-INDUCED GT	1.12e+01
29	59	32.1	900	3	FOX2_YEAST	PEROXISOMAL HYDRATASE	1.12e+01
30	59	32.1	767	6	TOP1_CRIGR	DNA TOPOISOMERASE I (1.12e+01
31	58	31.5	846	7	VAV_HUMAN	VAV ONCOGENE.	1.61e+01
32	58	31.5	333	7	XYNB_STRLI	ENDO-1,4-BETA-XYLANS	1.61e+01
33	58	31.5	423	5	PSY_ARATH	PHYTOENE SYNTHASE PRE	1.61e+01
34	58	31.5	550	5	PHNL_DESGI	PERIPLASMIC [NIFE] HY	1.61e+01
35	58	31.5	249	1	BA71_EUBSP	7-ALPHA-HYDROXYSTEROI	1.61e+01
36	58	31.5	986	2	EPIB_STAEP	EPIDERMIN BIOSYNTHESI	1.61e+01
37	58	31.5	845	7	VAV_MOUSE	VAV PROTO-ONCOGENE.	1.61e+01
38	58	31.5	431	4	K1CX_HUMAN	KERATIN, TYPE I CYTOS	1.61e+01
39	57	31.0	2150	6	SDC3_CAEEL	SDC-3 PROTEIN.	2.30e+01
40	57	31.0	1131	7	YA19_YEAST	HYPOTHETICAL 128.5 KD	2.30e+01
41	57	31.0	906	2	CTNA_HUMAN	ALPHA-CATENIN (CADHER	2.30e+01
42	57	31.0	698	2	CRAC_DICDI	PROTEIN CRAC.	2.30e+01
43	57	31.0	906	2	CTNA_MOUSE	ALPHA-CATENIN (102 KD	2.30e+01
44	57	31.0	116	6	RL21_MARPO	50S RIBOSOMAL PROTEIN	2.30e+01
45	57	31.0	192	3	HS41_SOYBN	22.0 KD CLASS IV HEAT	2.30e+01

ALIGNMENTS

RESULT 1
ID FELB_FELCA STANDARD; PRT; 88 AA.
AC P30439;
DT 01-APR-1993 (REL. 25, CREATED)
DT 01-APR-1993 (REL. 25, LAST SEQUENCE UPDATE)
DT 01-APR-1993 (REL. 25, LAST SEQUENCE UPDATE)

DE MAJOR ALLERGEN I POLYPEPTIDE CHAIN 1 MINOR FORM PRECURSOR (FEL D 1)
 DE (CAT-1) (AG 4).
 GN CH1.
 OS FELIS CATUS (CAT).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
 OC EUTHERIA; CARNIVORA.
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 19-88.
 RM 92052157
 RA MORGENSTERN J.P., GRIFFITH I.J., BRAUER A.W., ROGERS B.L.,
 RA BOND J.F., CHAPMAN M.D., KUO M.-C.;
 RL PROC. NATL. ACAD. SCI. U.S.A. 88:9690-9694(1991).
 RN [2]
 RP SEQUENCE FROM N.A.
 RM 92241678
 RA GRIFFITH I.J., CRAIG S., POLLOCK J., YU X.-B., MORGENSTERN J.P.,
 RA ROGERS B.L.,
 RL GENE 113:263-268(1992).
 RN [3]
 RP SEQUENCE OF 19-58, AND CHARACTERIZATION.
 RM 91287714
 RA DUFFORT D.A., CARREIRA J., NITTI G., POLO F., LOMBARDEO M.;
 RL MOL. IMMUNOL. 28:301-309(1991).
 RN [4]
 RP CHARACTERIZATION.
 RA LEITERMANN K., OHMAN J.L. JR.;
 RL J. ALLERGY CLIN. IMMUNOL. 74:147-153(1991).
 CC -!- DISEASE: MAJOR ALLERGEN PRODUCED BY THE DOMESTIC CAT.
 CC -!- SUBUNIT: HETEROTETRAMER COMPOSED OF TWO NON-COVALENTLY LINKED
 CC DISULFIDE-LINKED HETERODIMER OF CHAINS 1 AND 2.
 CC -!- TISSUE SPECIFICITY: SALIVA, AND SEBACEOUS GLANDS.
 CC -!- ALTERNATIVE PRODUCTS: USAGE OF TWO DIFFERENT INITIATOR MET ARE
 CC RESPONSIBLE FOR THE PRODUCTION OF TWO FORMS OF THE SIGNAL SEQUENCE
 CC OF THIS ALLERGEN SUBUNIT.
 CC -!- SIMILARITY: TO UTEROGLOBIN.
 DR EMBL; M74953; FDFELDIB.
 DR PIR; JC1126; JC1126.
 DR PROSITE; PS00403; UTEROGLOBIN_1.
 DR PROSITE; PS00404; UTEROGLOBIN_2.
 KW ALLERGEN; SIGNAL; ALTERNATIVE SPLICING.
 FT SIGNAL 1 18
 FT CHAIN 19 88 MAJOR ALLERGEN I POLYPEPTIDE CHAIN 1.
 FT DISULFID 21 21 INTERCHAIN (POTENTIAL).
 FT DISULFID 88 88 INTERCHAIN (POTENTIAL).
 FT VARIANT 47 47 K -> N.
 FT CONFLICT 78 78 L -> V (IN REF. 2).
 SQ SEQUENCE 88 AA; 9614 MW; 39445 CN;

DB 2; Score 184; Match 100.0%; Predicted No. 1.03e-27;
 Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 47 kalpvvlenarilkncvdakmteedke 73
 |||||
 Qy 1 KALPVVLENARILKNCVDKMTTEEDKE 27

RESULT 2
 ID FELA_FELCA STANDARD; PRT; 92 AA.
 AC P30438;
 DT 01-APR-1993 (REL. 25, CREATED)
 DT 01-APR-1993 (REL. 25, LAST SEQUENCE UPDATE)
 DT 01-JUN-1994 (REL. 29, LAST ANNOTATION UPDATE)
 DE MAJOR ALLERGEN I POLYPEPTIDE CHAIN 1 MAJOR FORM PRECURSOR (FEL D 1)
 DE (CAT-1) (AG 4).
 GN CH1.
 OS FELIS CATUS (CAT)

```

RESULT      3
ID   POLG_FMDV1      STANDARD;      PRT:  2333 AA.
AC   P03306;
DT   21-JUL-1986 (REL. 01, CREATED)
DT   21-JUL-1986 (REL. 01, LAST SEQUENCE UPDATE)
DT   01-OCT-1994 (REL. 30, LAST ANNOTATION UPDATE)
DE   GENOME POLYPROTEIN (NONSTRUCTURAL PROTEIN P20A; COAT PROTEINS VP1 TO
DE   VP4; CORE PROTEIN P52; GENOME-LINKED PROTEINS VPG1 TO VPG3; PICORNAIN
DE   3C (EC 3.4.22.28) (PROTEASE 3C) (P3C); RNA-DIRECTED RNA POLYMERASE
DE   (EC 2.7.7.48)).
CD   FOOT-AND-MOUTH DISEASE VIRUS (STRAIN A10-11) (APHTHOVIRUS A)

```

DC VIRIDAE; SS-RNA NONENVELOPED VIRUSES; PICORNAVIRIDAE; APHTHOVIRUSES.
 RN [1]
 RP SEQUENCE FROM N.A.
 RM 84169547
 RA CARROLL A.R., ROWLANDS D.J., CLARKE B.E.;
 RL NUCLEIC ACIDS RES. 12:2461-2472(1984).
 RN [2]
 RP SEQUENCE OF 115-1048 FROM N.A.
 RM 82211814
 RA BOOTHROYD J.C., HARRIS T.J.R., ROWLANDS D.J., LOWE P.A.;
 RL GENE 17:153-161(1982).
 CC -!- PTM: SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
 CC -!- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAHEDRAL UNITS,
 CC EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,
 CC VP3, AND VP4.
 DR EMBL; X00429; PIFMDV1.
 DR PIR; A03908; GNNY2F.
 KW POLYPROTEIN; COAT PROTEIN; CORE PROTEIN; RNA-DIRECTED RNA POLYMERASE;
 KW HYDROLASE; THIOL PROTEASE; MYRISTYLATION.
 FT CHAIN 1 201 NONSTRUCTURAL PROTEIN P20A.
 FT CHAIN 202 286 COAT PROTEIN VP4.
 FT CHAIN 287 504 COAT PROTEIN VP2.
 FT CHAIN 505 725 COAT PROTEIN VP3.
 FT CHAIN 726 937 COAT PROTEIN VP1.
 FT CHAIN 938 1578 CORE PROTEIN P52.
 FT CHAIN 1579 1601 GENOME-LINKED PROTEIN VPG1.
 FT CHAIN 1602 1625 GENOME-LINKED PROTEIN VPG2.
 FT CHAIN 1626 1649 GENOME-LINKED PROTEIN VPG3.
 FT CHAIN 1650 1863 PROTEASE P20B.
 FT CHAIN 1864 2333 RNA-DIRECTED RNA POLYMERASE P56A.
 FT LIPID 202 202 MYRISTATE.
 FT CONFLICT 396 396 S -> C (IN REF. 2).
 FT CONFLICT 632 632 P -> L (IN REF. 2).
 SQ SEQUENCE 2333 AA; 259645 MW; 19388774 CN;

DB 5; Score 79; Match 59.1%; Predicted No. 3.09e-03;
 Matches 13; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

Db 1913 vvlddvifskhkgdakmteedk 1934
 |||:: : | : |||||
 Qy 5 VVLENARILKNCVDAKMTEEDK 26

RESULT 4
 ID POLG_FMDVO STANDARD; PRT; 2332 AA.
 AC P03305;
 DT 21-JUL-1986 (REL. 01, CREATED)
 DT 21-JUL-1986 (REL. 01, LAST SEQUENCE UPDATE)
 DT 01-OCT-1994 (REL. 30, LAST ANNOTATION UPDATE)
 DE GENOME POLYPROTEIN (NONSTRUCTURAL PROTEIN P20A; COAT PROTEINS VP1 TO
 DE VP4; CORE PROTEINS P12, P34, P14; GENOME-LINKED PROTEIN VPG; PROTEASE
 DE (EC 3.4.22.-); RNA-DIRECTED RNA POLYMERASE (EC 2.7.7.48)).
 OS FOOT-AND-MOUTH DISEASE VIRUS (STRAINS O1K AND O1BFS) (APHTHOVIRUS O).
 DC VIRIDAE; SS-RNA NONENVELOPED VIRUSES; PICORNAVIRIDAE; APHTHOVIRUSES.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=O1K;
 RM 84297249
 RA FORSS S., STREBEL K., BECK E., SCHALLER H.;
 RL NUCLEIC ACIDS RES. 12:6587-6601(1984).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=O1BFS;
 RM 83143292
 RA MAKOFF A.J., PAYNTER C.A., ROWLANDS D.J., BOOTHROYD J.C.;
 NUCLEIC ACIDS RES. 12:6587-6601(1984)

RN [3]
 RP X-RAY CRYSTALLOGRAPHY (2.9 ANGSTROMS).
 RM 89143740
 RA ACHARYA R., FRY E., STUART D., FOX G., ROWLANDS D., BROWN F.;
 RL NATURE 337:709-716(1989).
 CC -!- THE STRAIN 01K SEQUENCE IS SHOWN.
 CC -!- PTM: SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
 CC -!- THE COAT PROTEIN VP1 CONTAINS THE MAIN ANTIGENIC DETERMINANTS OF
 CC THE VIRION; THEREFORE, CHANGES IN ITS SEQUENCE MUST BE
 CC RESPONSIBLE FOR THE HIGH ANTIGENIC VARIABILITY OF THE VIRUS.
 CC -!- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAEDRAL UNITS,
 CC EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,
 CC VP3, AND VP4.
 DR EMBL; X00871; PIFMDV2.
 DR EMBL; J02185; P101VP.
 DR PIR; A03907; GNNYF.
 KW POLYPROTEIN; COAT PROTEIN; CORE PROTEIN; RNA-DIRECTED RNA POLYMERASE;
 KW HYDROLASE; THIOL PROTEASE; MYRISTYLATION.
 FT CHAIN 1 201 NONSTRUCTURAL PROTEIN P20A.
 FT CHAIN 202 286 COAT PROTEIN VP4.
 FT CHAIN 287 504 COAT PROTEIN VP2.
 FT CHAIN 505 724 COAT PROTEIN VP3.
 FT CHAIN 725 937 COAT PROTEIN VP1.
 FT CHAIN 938 1107 CORE PROTEIN P12.
 FT CHAIN 1108 1425 CORE PROTEIN P34.
 FT CHAIN 1426 1578 CORE PROTEIN P14.
 FT CHAIN 1579 1649 GENOME-LINKED PROTEIN VPG.
 FT CHAIN 1650 1862 PROTEASE.
 FT CHAIN 1863 2332 RNA-DIRECTED RNA POLYMERASE.
 FT LIPID 202 202 MYRISTATE.
 FT DISULFID 511 511 INTERCHAIN (IN VP3 DIMER).
 FT DISULFID 406 858 IN VP2-VP1 DIMER.
 FT VARIANT 780 780 I -> V (IN STRAIN 01BFS).
 FT VARIANT 808 808 G -> R (IN STRAIN 01BFS).
 FT VARIANT 861 861 N -> S (IN STRAIN 01BFS).
 SQ SEQUENCE 2332 AA; 258924 MW; 19411374 CN;

DB 5; Score 72; Match 50.0%; Predicted No. 6.49e-02;
 Matches 11; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

Db 1912 vvldevifskhkgtkmseedk 1933
 |||:: : | : |||:||||
 Qy 5 VVLENARILKNCVDAKMTEEDK 26

RESULT 5
 ID YLD1_CAEEL STANDARD; PRT: 374 AA.
 AC Q03566;
 DT 01-FEB-1994 (REL. 28, CREATED)
 DT 01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)
 DT 01-JUN-1994 (REL. 29, LAST ANNOTATION UPDATE)
 DE PROBABLE G PROTEIN-COUPLED RECEPTOR C38C10.1 IN CHROMOSOME III.
 GN C38C10.1.
 OS CAENORHABDITIS ELEGANS.
 OC EUKARYOTA; METAZOA; ACOELEMATES; NEMATODA; SECERNENTEA; RHABDITIDA.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BRISTOL N2;
 RM 94150718
 RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
 RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
 RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FRASER A.,
 RA FULTON L., GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M.,
 RA JOHNSTON L., JONES M., KERSHAW J., KIRSTEN J., LAISSTER N.,
 RA LATREILLE P., LIGHTNING J., LLOYD C., MORTIMORE B., O'CALLAGHAN M.,
 RA PARRIS B., PERCY J., REYEN B., ROOPRA A., SANDERS D., SHANKEN B.

RA SIMS M., SMALDON N., SMITH A., SMITH M., SONNHAMMER E., STADEN R.,
 RA SULSTON J., THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K.,
 RA WATERSON R., WATSON A., WEINSTOCK L., WILKINSON-SPROAT J.,
 RA WOHLDMAN P.;
 RL NATURE 368:32-38(1994).
 CC -!- FUNCTION: NOT KNOWN. PUTATIVE RECEPTOR.
 CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
 CC -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
 CC MOST SIMILAR TO TACHYKININS RECEPTORS.
 DR EMBL; Z19153; CEC38C10.
 DR PIR; S28285; S28285.
 DR WORMPEP; C38C10.1; CE00104.
 DR GCRDB; GCR_0567; -.
 DR PROSITE; PS00237; G_PROTEIN_RECEPTOR.
 KW HYPOTHETICAL PROTEIN; G-PROTEIN COUPLED RECEPTOR; TRANSMEMBRANE;
 KW GLYCOPROTEIN.
 SQ SEQUENCE 374 AA; 42940 MW; 769122 CN;

DB 7; Score 71; Match 22.2%; Predicted No. 9.89e-02;
 Matches 6; Conservative 14; Mismatches 7; Indels 0; Gaps 0;

Db 322 rsmailsqkgrvnsscldkkvknsq 348
 : : : : | : : : | : : | | : : :
 Qy 1 KALPVVLENARILKNCVDAKMTEEDKE 27

RESULT 6
 ID YCTP_ASTLO STANDARD; PRT; 136 AA.
 AC P34776;
 DT 01-FEB-1994 (REL. 28, CREATED)
 DT 01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)
 DT 01-FEB-1994 (REL. 28, LAST ANNOTATION UPDATE)
 DE HYPOTHETICAL PROTEIN IN TRNP 5'REGION (FRAGMENT).
 OS ASTASIA LONGA (EUGLENOPHYCEAN ALGA).
 OG CHLOROPLAST.
 OC EUKARYOTA; PLANTA; PHYCOPHYTA; EUGLENOPHYTA.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CCAP 1204-17A;
 RA GOCKEL G., BAIER S., HACHTEL W.;
 RL SUBMITTED (NOV-1993) TO EMBL/GENBANK/DBJ DATA BANKS.
 DR EMBL; X75653; ALRIBPTR.
 DR PIR; S38598; S38598.
 KW CHLOROPLAST; HYPOTHETICAL PROTEIN.
 FT NON_TER 1 1
 SQ SEQUENCE 136 AA; 16587 MW; 103277 CN;

DB 7; Score 71; Match 46.7%; Predicted No. 9.89e-02;
 Matches 7; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

Db 120 ldddrilnvcvitrn 134
 | : : | | : : | : : | : : | : :
 Qy 7 LENARILKNCVDAKM 21

RESULT 7
 ID LDHH_PIG STANDARD; PRT; 333 AA.
 AC P00336;
 DT 21-JUL-1986 (REL. 01, CREATED)
 DT 21-JUL-1986 (REL. 01, LAST SEQUENCE UPDATE)
 DT 01-MAR-1992 (REL. 21, LAST ANNOTATION UPDATE)
 DE L-LACTATE DEHYDROGENASE H CHAIN (EC 1.1.1.27) (LDH-B).
 OS SUS SCROFA (PIG).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
 OC EUTHERIA; ARTIODACTYLA.

RP SEQUENCE.
RM 77117453
RA KILTZ H.-H., KEIL W., GRIESBACH M., PETRY K., MEYER H.;
RL HOPPE-SEYLER'S Z. PHYSIOL. CHEM. 358:123-127(1977).
RN [2]
RP REVISIONS TO 21; 147; 215 AND 217.
RA KILTZ H.-H.;
RL SUBMITTED (OCT-1977) TO THE PIR DATA BANK.
RN [3]
RP X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS).
RM 82170431
RA GRAU U.M., TROMMER W.E., ROSSMANN M.G.;
RL J. MOL. BIOL. 151:289-307(1981).
CC -!- CATALYTIC ACTIVITY: L-LACTATE + NAD(+) = PYRUVATE + NADH.
CC -!- SUBUNIT: HOMOTETRAMER.
CC -!- PATHWAY: FINAL STEP IN ANAEROBIC GLYCOLYSIS.
CC -!- THERE ARE THREE TYPES OF LDH CHAINS: M (LDH-A) FOUND PREDOMINANTLY
CC IN MUSCLE TISSUES, H (LDH-B) FOUND IN HEART MUSCLE AND X (LDH-C)
CC WHICH IS PRESENT ONLY IN THE SPERMATOZOA OF MAMMALS AND BIRDS.
DR PIR; A00345; DEPLH.
DR PDB; 5LDH; 16-APR-88.
DR PROSITE; PS00064; L_LDH.
KW OXIDOREDUCTASE; NAD; GLYCOLYSIS; MULTIGENE FAMILY; ACETYLATION;
KW 3D-STRUCTURE.
FT MOD_RES 1 1 ACETYLATION.
FT ACT_SITE 193 193 ACCEPTS A PROTON DURING CATALYSIS.
FT STRAND 23 25
FT HELIX 30 42
FT STRAND 48 50
FT HELIX 56 71
FT STRAND 77 78
FT HELIX 85 87
FT STRAND 91 94
FT TURN 97 98
FT TURN 106 107
FT HELIX 111 118
FT TURN 119 120
FT HELIX 121 127
FT STRAND 132 135
FT HELIX 140 151
FT HELIX 154 157
FT STRAND 158 159
FT TURN 162 163
FT HELIX 164 178
FT STRAND 186 186
FT STRAND 189 190
FT STRAND 198 199
FT STRAND 205 205
FT TURN 214 217
FT HELIX 227 239
FT HELIX 246 265
FT TURN 266 266
FT STRAND 271 271
FT STRAND 293 296
FT TURN 297 298
FT STRAND 299 303
FT HELIX 312 330
SQ SEQUENCE 333 AA; 36476 MW; 600619 CN;

DB 4; Score 68; Match 37.5%; Predicted No. 3.42e-01;
Matches 9; Conservative 8; Mismatches 6; Indels 1; Gaps 1;

Db 290 slpcvl-nargltsvinqklkdde 312
:||||| ||| | : :: |: ::
Qy 2 ALPVVLENARILKNCVDAKMTED 25

RESULT 8

ID LDHH_HUMAN STANDARD; PRT; 333 AA.
AC P07195;
DT 01-APR-1988 (REL. 07, CREATED)
DT 01-APR-1988 (REL. 07, LAST SEQUENCE UPDATE)
DT 01-JUN-1994 (REL. 29, LAST ANNOTATION UPDATE)
DE L-LACTATE DEHYDROGENASE H CHAIN (EC 1.1.1.27) (LDH-B).
GN LDHB.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; PRIMATES.
RN [1]
RP SEQUENCE FROM N.A.
RM 89193506
RA TAKENO T., LI S.S.-L.;
RL BIOCHEM. J. 257:921-924(1989).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=T-CELL;
RM 88133965
RA SAKAI I., SHARIEF F.S., PAN Y.-C.E., LI S.S.-L.;
RL BIOCHEM. J. 248:933-936(1987).
RN [3]
RP VARIANT GLU-6.
RM 93300477
RA MAEKAWA M., SUDO K., KITAJIMA M., MATSUURA Y., LI S.S.-L., KANNO T.;
RL HUM. GENET. 91:423-426(1993).
RN [4]
RP VARIANTS GLU-34; VAL-170 AND LEU-174.
RM 93216283
RA MAEKAWA M., SUDO K., KITAJIMA M., MATSUURA Y., LI S.S.-L., KANNO T.;
RL HUM. GENET. 91:163-168(1993).
RN [5]
RP VARIANTS ARG-128 AND HIS-171.
RM 92267519
RA SUDO K., MAEKAWA M., TOMONAGA A., TSUKADA T., NAKAYAMA T.,
RA KITAMURA M., LI S.S.-L., KANNO T., TORIUMI J.;
RL HUM. GENET. 89:158-162(1992).
RN [6]
RP VARIANT HIS-171.
RM 90241237
RA SUDO K., MAEKAWA M., IKAWA S., MACHIDA K., KITAMURA M., LI S.S.-L.;
RL BIOCHEM. BIOPHYS. RES. COMMUN. 168:672-676(1990).
CC -!- CATALYTIC ACTIVITY: L-LACTATE + NAD(+) = PYRUVATE + NADH.
CC -!- SUBUNIT: HOMOTETRAMER.
CC -!- PATHWAY: FINAL STEP IN ANAEROBIC GLYCOLYSIS.
CC -!- THERE ARE THREE TYPES OF LDH CHAINS: M (LDH-A) FOUND PREDOMINANTLY
CC IN MUSCLE TISSUES, H (LDH-B) FOUND IN HEART MUSCLE AND X (LDH-C)
CC WHICH IS PRESENT ONLY IN THE SPERMATOZOA OF MAMMALS AND BIRDS.
CC -!- DISEASE: LDHB DEFICIENCY PROBABLY HAS NO CLEAR SYMPTOMATIC
CC CONSEQUENCES.
DR EMBL; Y00711; HSLDHB1.
DR EMBL; X13794; HSLDHB1.
DR EMBL; X13795; HSLDHB3.
DR EMBL; X13796; HSLDHB4.
DR EMBL; X13797; HSLDHB5.
DR EMBL; X13798; HSLDHB6.
DR EMBL; X13799; HSLDHB7.
DR EMBL; X13800; HSLDHB8.
DR PIR; S02795; S02795.
DR NIM; 150100; 11TH EDITION.
DR PROSITE; PS00064; L_LDH.
KW OXIDOREDUCTASE; NAD; GLYCOLYSIS; MULTIGENE FAMILY; DISEASE MUTATION.
FT INIT_MET 0 0
FT 107 SITE 107 107 ACCEPTS A PROTON DURING CATALYSIS

FT	VARIANT	6	6	K -> E (IN LDHB DEFICIENCY; SLIGHTLY DECREASED ACTIVITY).
FT				
FT	VARIANT	34	34	A -> E (IN LDHB DEFICIENCY).
FT	VARIANT	128	128	S -> R (IN LDHB DEFICIENCY).
FT	VARIANT	170	170	F -> V (IN LDHB DEFICIENCY).
FT	VARIANT	171	171	R -> H (IN LDHB DEFICIENCY; UNSTABLE).
FT	VARIANT	174	174	M -> L (IN LDHB DEFICIENCY).
SQ	SEQUENCE	333 AA; 36507 MW; 598763 CN;		

DB 4; Score 66; Match 33.3%; Predicted No. 7.66e-01;
Matches 8; Conservative 9; Mismatches 6; Indels 1; Gaps 1;

Db 290 slpcil-nargltsvinqklkdde 312
: || : | ||| | : : : | : : :
Qy 2 ALPVVLENARILKNCVDAKMTEED 25

RESULT 9

ID POLG_FMDVA STANDARD; PRT: 2332 AA.
AC P03308; P03312;
DT 21-JUL-1986 (REL. 01, CREATED)
DT 01-JAN-1988 (REL. 06, LAST SEQUENCE UPDATE)
DT 01-OCT-1994 (REL. 30, LAST ANNOTATION UPDATE)
DE GENOME POLYPROTEIN (NONSTRUCTURAL PROTEIN P20A; COAT PROTEINS VP1 TO
DE VP4; CORE PROTEINS X, P14, P41, P19; GENOME-LINKED PROTEINS VPG1 TO
DE VPG3; PICORNAIN 3C (EC 3.4.22.28) (PROTEASE 3C) (P3C); RNA-DIRECTED
DE RNA POLYMERASE (EC 2.7.7.48)).
OS FOOT-AND-MOUTH DISEASE VIRUS (STRAIN A12) (APHTHOVIRUS A).
OC VIRIDAE; SS-RNA NONENVELOPED VIRUSES; PICORNAVIRIDAE; APHTHOVIRUSES.
RN [1]
RP SEQUENCE FROM N.A.
RM 85211015
RA ROBERTSON B.H., GRUBMAN M.J., WEDDELL G.N., MOORE D.M., WELSH J.D.,
RA FISCHER T., DOWBENKO D.J., YANSURA D.G., SMALL B., KLEID D.G.;
RL J. VIROL. 54:651-660(1985).
RN [2]
RP SEQUENCE OF 1863-2332 FROM N.A.
RM 83225613
RA ROBERTSON B.H., MORGAN D.O., MOORE D.M., GRUBMAN M.J., CARD J.,
RA FISCHER T., WEDDELL G.N., DOWBENKO D.J., YANSURA D.G.;
RL VIROLOGY 126:614-623(1983).
RN [3]
RP SEQUENCE OF 715-955 FROM N.A.
RM 82061853
RA KLEID D.G., YANSURA D.G., SMALL B., DOWBENKO D.J., MOORE D.M.,
RA GRUBMAN M.J., MCKERCHER P.D., MORGAN D.O., ROBERTSON B.H.,
RA BACHRACH H.L.;
RL SCIENCE 214:1125-1129(1981).
CC -!- PTM: SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
CC -!- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAHEDRAL UNITS,
CC EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,
CC VP3, AND VP4.
DR EMBL; M10975; APHA12CD.
DR PIR; A25794; GNNY4F.
KW POLYPROTEIN; COAT PROTEIN; CORE PROTEIN; RNA-DIRECTED RNA POLYMERASE;
KW HYDROLASE; THIOL PROTEASE; MYRISTYLATION.
FT CHAIN 1 200 NONSTRUCTURAL PROTEIN P20A.
FT CHAIN 201 285 COAT PROTEIN VP4.
FT CHAIN 286 503 COAT PROTEIN VP2.
FT CHAIN 504 723 COAT PROTEIN VP3.
FT CHAIN 724 937 COAT PROTEIN VP1.
FT CHAIN 938 953 CORE PROTEIN X.
FT CHAIN 954 1107 CORE PROTEIN P14.
FT CHAIN 1108 1425 CORE PROTEIN P41.
FT CHAIN 1426 1578 CORE PROTEIN P19.
FT CHAIN 1579 1601 GENOME-LINKED PROTEIN VPG1

FT CHAIN 1602 1625 GENOME-LINKED PROTEIN VPG2.
 FT CHAIN 1626 1649 GENOME-LINKED PROTEIN VPG3.
 FT CHAIN 1650 1862 PROTEASE.
 FT CHAIN 1863 2332 RNA-DIRECTED RNA POLYMERASE.
 FT LIPID 201 201 MYRISTATE.
 SQ SEQUENCE 2332 AA; 259408 MW; 19347576 CN;

DB 5; Score 66; Match 45.5%; Predicted No. 7.66e-01;
 Matches 10; Conservative 6; Mismatches 6; Indels 0; Gaps 0;

Db 1912 vvlddevifskhkgtkmsaedk 1933
 |||: : | : ||| |||
 Qy 5 VVLENARILKNCVDAKMTEEDK 26

RESULT 10

ID LDHH_RABIT STANDARD; PRT: 217 AA.
 AC P13490;
 DT 01-JAN-1990 (REL. 13, CREATED)
 DT 01-JAN-1990 (REL. 13, LAST SEQUENCE UPDATE)
 DT 01-JAN-1990 (REL. 13, LAST ANNOTATION UPDATE)
 DE L-LACTATE DEHYDROGENASE H CHAIN (EC 1.1.1.27) (LDH-B) (FRAGMENT).
 OS ORYCTOLAGUS CUNICULUS (RABBIT).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
 OC EUTHERIA; LAGOMORPHA.
 RN [1]
 RP SEQUENCE FROM N.A.
 RM 89139477
 RA SASS C., BRIAND M., BENSLIMANE S., RENAUD M., BRIAND Y.;
 RL J. BIOL. CHEM. 264:4076-4081(1989).
 CC -!- CATALYTIC ACTIVITY: L-LACTATE + NAD(+) = PYRUVATE + NADH.
 CC -!- SUBUNIT: HOMOTETRAMER.
 CC -!- PATHWAY: FINAL STEP IN ANAEROBIC GLYCOLYSIS.
 CC -!- THERE ARE THREE TYPES OF LDH CHAINS: M (LDH-A) FOUND PREDOMINANTLY
 CC IN MUSCLE TISSUES, H (LDH-B) FOUND IN HEART MUSCLE AND X (LDH-C)
 CC WHICH IS PRESENT ONLY IN THE SPERMATOZOA OF MAMMALS AND BIRDS.
 DR EMBL; M22584; OCLDHH.
 DR PIR; B32957; B32957.
 DR PROSITE; PS00064; L_LDH.
 KW OXIDOREDUCTASE; NAD; GLYCOLYSIS; MULTIGENE FAMILY.
 FT NON_TER 1 1
 FT ACT_SITE 77 77 ACCEPTS A PROTON DURING CATALYSIS.
 SQ SEQUENCE 217 AA; 24134 MW; 249993 CN;

DB 4; Score 66; Match 33.3%; Predicted No. 7.66e-01;
 Matches 8; Conservative 9; Mismatches 6; Indels 1; Gaps 1;

Db 174 slpcil-nargltsvinqklkdde 196
 :|| :| ||| | : :: | : ::
 Qy 2 ALPVVLENARILKNCVDAKMTEED 25

RESULT 11

ID LDHH_MOUSE STANDARD; PRT: 333 AA.
 AC P16125;
 DT 01-APR-1990 (REL. 14, CREATED)
 DT 01-APR-1990 (REL. 14, LAST SEQUENCE UPDATE)
 DT 01-AUG-1990 (REL. 15, LAST ANNOTATION UPDATE)
 DE L-LACTATE DEHYDROGENASE H CHAIN (EC 1.1.1.27) (LDH-B).
 GN LDH-2.
 OS MUS MUSCULUS (MOUSE).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
 OC EUTHERIA; RODENTIA.
 RN [1]
 RP SEQUENCE FROM N.A.
 PM 89008712

RA HIRAOKA B.Y., SHARIEF F.S., YANG Y.W., LI W.H., LI S.S.-L.;
 RL EUR. J. BIOCHEM. 189:215-220(1990).
 CC -!- CATALYTIC ACTIVITY: L-LACTATE + NAD(+) = PYRUVATE + NADH.
 CC -!- SUBUNIT: HOMOTETRAMER.
 CC -!- PATHWAY: FINAL STEP IN ANAEROBIC GLYCOLYSIS.
 CC -!- THERE ARE THREE TYPES OF LDH CHAINS: M (LDH-A) FOUND PREDOMINANTLY
 CC IN MUSCLE TISSUES, H (LDH-B) FOUND IN HEART MUSCLE AND X (LDH-C)
 CC WHICH IS PRESENT ONLY IN THE SPERMATOZOA OF MAMMALS AND BIRDS.
 DR EMBL; X51905; MMLDH2.
 DR PIR; S09954; S09954.
 DR PROSITE; PS00064; L_LDH.
 KW OXIDOREDUCTASE; NAD; GLYCOLYSIS; MULTIGENE FAMILY.
 FT INIT_MET 0 0
 FT ACT_SITE 193 193 ACCEPTS A PROTON DURING CATALYSIS.
 SQ SEQUENCE 333 AA; 36441 MW; 595466 CN;

DB 4; Score 66; Match 33.3%; Predicted No. 7.66e-01;
 Matches 8; Conservative 9; Mismatches 6; Indels 1; Gaps 1;

Db 290 slpcil-nargltsvinqklkdde 312
 :||:| ||| | : :: | : ::
 Qy 2 ALPVVLENARILKNCVDAKMTED 25

RESULT 12

ID MYSD_CHICK STANDARD; PRT; 1829 AA.
 AC Q02440;
 DT 01-JUN-1994 (REL. 29, CREATED)
 DT 01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)
 DT 01-OCT-1994 (REL. 30, LAST ANNOTATION UPDATE)
 DE DILUTE MYOSIN HEAVY CHAIN, ISOFORM I.
 OS GALLUS GALLUS (CHICKEN).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; AVES; NEOGNATHAE;
 OC GALLIFORMES.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=BRAIN;
 RM 93012002
 RA SANDERS G., LICHT E B., MEYER H.E., KILIMANN M.W.;
 RL FEBS LETT. 311:295-298(1992).
 CC -!- SIMILARITY: BELONGS TO CLASS-5 MYOSINS.
 DR EMBL; X67251; GGDILUTE.
 KW MYOSIN; REPEAT; ATP-BINDING; CALMODULIN-BINDING;
 KW HEPTAD REPEAT PATTERN.
 FT NP_BIND 163 170 ATP (BY SIMILARITY).
 SQ SEQUENCE 1829 AA; 212381 MW; 15626072 CN;

DB 4; Score 66; Match 39.1%; Predicted No. 7.66e-01;
 Matches 9; Conservative 6; Mismatches 8; Indels 0; Gaps 0;

Db 36 kvlqlrleegkdleycldpktke 58
 | | : ||::: | |::| |
 Qy 1 KALPVVLENARILKNCVDAKMT E 23

RESULT 13

ID ACRR_ECOLI STANDARD; PRT; 215 AA.
 AC P34000;
 DT 01-FEB-1994 (REL. 28, CREATED)
 DT 01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)
 DT 01-OCT-1994 (REL. 30, LAST ANNOTATION UPDATE)
 DE POTENTIAL REPRESSOR FOR ACRA B OPERON.
 GN ACRR.
 OS ESCHERICHIA COLI.
 OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; FACULTATIVELY ANAEROBIC RODS;
 CC ENTEROBACTERIACEAE

RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / W4573;
 RM 94012493
 RA MA D., COOK D.N., ALBERTI M., PON N.G., NIKAIDO H., HEARST J.E.;
 RL J. BACTERIOL. 175:6299-6313(1993).
 RN [2]
 RP IDENTIFICATION.
 RA RUDD K.E.;
 RL UNPUBLISHED OBSERVATIONS (DEC-1993).
 CC -!- FUNCTION: POTENTIAL REGULATOR PROTEIN FOR THE ACRA B GENES.
 CC -!- SIMILARITY: BELONGS TO THE ACRR/TTK FAMILY.
 DR EMBL; U00734; EC734.
 DR ECGENE; EG12116; ACRR.
 KW TRANSCRIPTION REGULATION; DNA-BINDING; REPRESSOR.
 SQ SEQUENCE 215 AA; 24766 MW; 223841 CN;

DB 1; Score 64; Match 46.7%; Predicted No. 1.69e+00;
 Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Db 143 qtlkhcieakmlpad 157
 : ||:|:| |
 Qy 11 RILKNCVDAKMTEED 25

RESULT 14

ID HDE_CANTR STANDARD; PRT; 906 AA.
 AC P22414;
 DT 01-AUG-1991 (REL. 19, CREATED)
 DT 01-AUG-1991 (REL. 19, LAST SEQUENCE UPDATE)
 DT 01-DEC-1992 (REL. 24, LAST ANNOTATION UPDATE)
 DE HYDRATASE-DEHYDROGENASE-EPIMERASE (HDE).
 OS CANDIDA TROPICALIS (YEAST).
 OC EUKARYOTA; FUNGI; DEUTEROMYCOTINA (IMPERFECT FUNGI).
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 20336 / PK233;
 RM 89172062
 RA NUTTLEY W.M., AITCHISON J.D., RACHUBINSKI R.A.;
 RL GENE 69:171-180(1988).
 RN [2]
 RP SIMILARITY TO SHORT CHAIN DEHYDROGENASES OF N-TERMINAL DOMAIN.
 RM 90367890
 RA BAKER M.E.;
 RL FASEB J. 4:3028-3032(1990).
 CC -!- FUNCTION: SECOND TRIFUNCTIONAL ENZYME ACTING ON THE BETA-OXIDATION
 CC PATHWAY FOR FATTY ACIDS, POSSESSING HYDRATASE-DEHYDROGENASE-
 CC EPIMERASE ACTIVITIES.
 CC -!- PATHWAY: BETA-OXIDATION PATHWAY.
 CC -!- INDUCTION: BY GROWTH ON N-ALKANES OR FATTY ACIDS.
 CC -!- SUBUNIT: MONOMER.
 CC -!- SUBCELLULAR LOCATION: PEROXISOMAL.
 CC -!- SIMILARITY: THE N-TERMINAL PART CONTAINS TWO COPIES OF INSECT-TYPE
 CC ALCOHOL DEHYDROGENASE / RIBITOL DEHYDROGENASE FAMILY DOMAIN.
 DR EMBL; M22765; M22765.
 DR PIR; JT0350; JT0350.
 DR PROSITE; PS00061; ADH_SHORT.
 DR PROSITE; PS00342; MICROBODIES_CTER.
 KW FATTY ACID METABOLISM; MULTIFUNCTIONAL ENZYME; OXIDOREDUCTASE; NAD;
 KW LYASE; ISOMERASE; PEROXISOME; DUPLICATION.
 FT DOMAIN 5 228 SHORT-CHAIN DEHYDROGENASE LIKE.
 FT DOMAIN 319 532 SHORT-CHAIN DEHYDROGENASE LIKE.
 FT SITE 904 906 MICROBODY TARGETING SIGNAL (POTENTIAL).
 SQ SEQUENCE 906 AA; 99409 MW; 4146036 CN;

Matches 10; Conservative 7; Mismatches 7; Indels 0; Gaps 0;

Db 91 tvhviinnagilrdasmkkmtkd 114
:: |:::| |:: |:: |
Qy 2 ALPVVLENARILKNCVDAKMTEED 25

RESULT 15

ID POLG_FMDVS STANDARD; PRT; 861 AA.
AC P03311;
DT 21-JUL-1986 (REL. 01, CREATED)
DT 01-NOV-1988 (REL. 09, LAST SEQUENCE UPDATE)
DT 01-OCT-1994 (REL. 30, LAST ANNOTATION UPDATE)
DE GENOME POLYPROTEIN (COAT PROTEINS VP3, VP1; CORE PROTEIN P52, PROTEASE
DE (EC 3.4.22.-); RNA-DIRECTED RNA POLYMERASE (EC 2.7.7.48)) (FRAGMENTS).
OS FOOT-AND-MOUTH DISEASE VIRUS (STRAIN C1-SANTA PAU [C-S8]) (APHTHOVIRUS
OS C).
OC VIRIDAE; SS-RNA NONENVELOPED VIRUSES; PICORNAVIRIDAE; APHTHOVIRUSES.
RN [1]
RP SEQUENCE OF 1-332 FROM N.A.
RM 84005890
RA VILLANUEVA N., DAVILA M., ORTIN J., DOMINGO E.;
RL GENE 23:185-194(1983).
RN [2]
RP SEQUENCE OF 333-861 FROM N.A.
RM 85286357
RA MARTINEZ-SALAS E., ORTIN J., DOMINGO E.;
RL GENE 35:55-61(1985).
CC -!- PTM: SPECIFIC ENZYMOLOGICAL CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
CC -!- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAHEDRAL UNITS,
CC EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,
CC VP3, AND VP4.
DR EMBL; M11027; PIP61.
DR PIR; A03913; A03913.
DR PIR; A24031; A24031.
KW POLYPROTEIN; COAT PROTEIN; CORE PROTEIN; RNA-DIRECTED RNA POLYMERASE;
KW HYDROLASE; THIOL PROTEASE.
FT NON_TER 1 1
FT CHAIN <1 46 COAT PROTEIN VP3.
FT CHAIN 47 254 COAT PROTEIN VP1.
FT CHAIN 255 332 CORE PROTEIN P52.
FT NON_CONS 332 333
FT CHAIN <333 391 PROTEASE.
FT CHAIN 392 861 RNA-DEPENDENT RNA POLYMERASE.
SQ SEQUENCE 861 AA; 95554 MW; 3818070 CN;

DB 5; Score 63; Match 40.9%; Predicted No. 2.49e+00;
Matches 9; Conservative 7; Mismatches 6; Indels 0; Gaps 0;

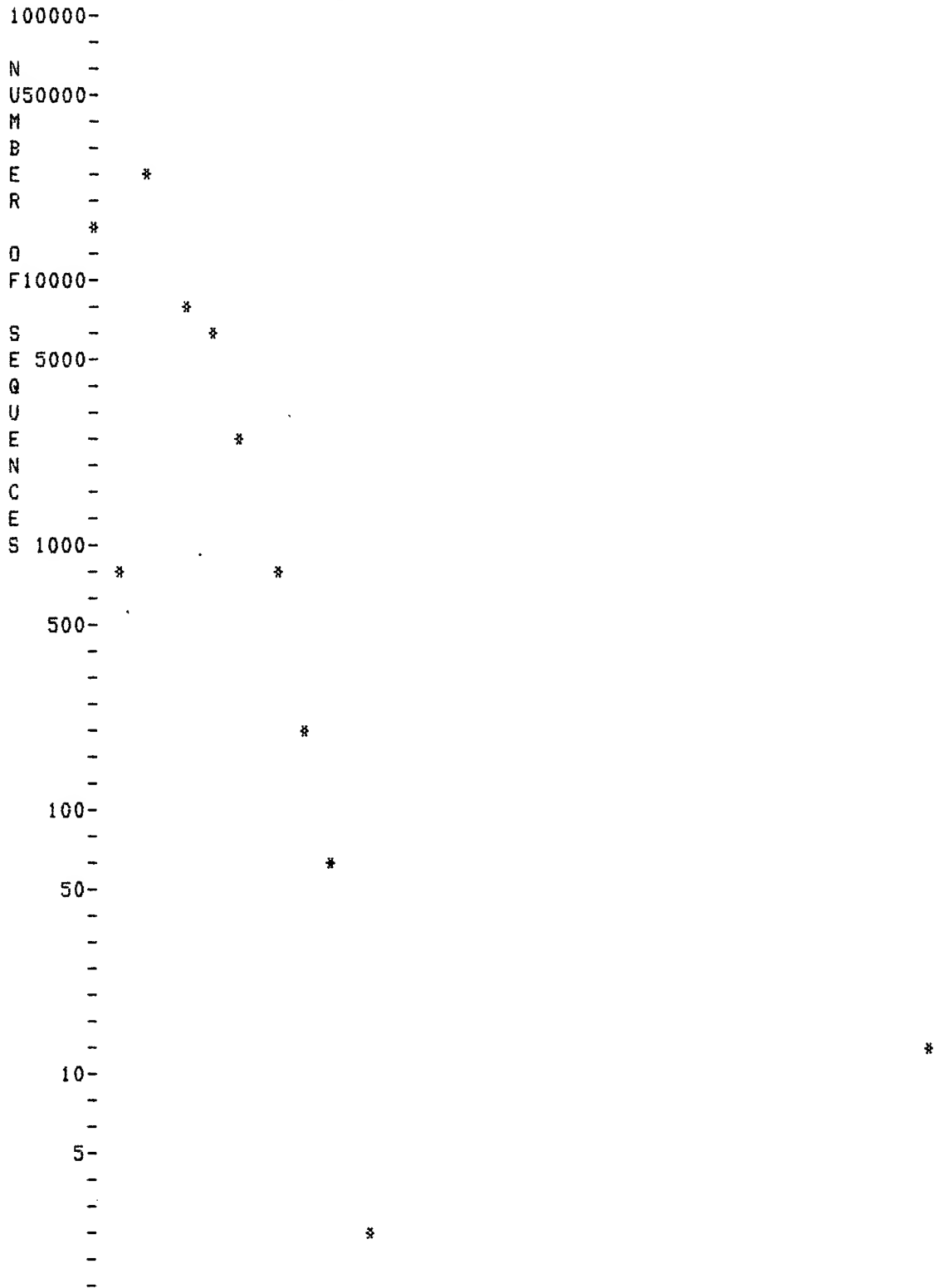
Db 441 vvldevifsrhkgdtkmsaedk 462
|:::| : : |:::| |
Qy 5 VVLENARILKNCVDAKMTEEDK 26

Search completed: Fri Mar 24 07:43:00 1995
Job time : 12 secs.

300,510

Results file 1-pat.res made by on Fri 24 Mar 95 7:55:32-PST.

Results of the initial comparison of US-08-300-510-1 (1-27) with:
Data bank : A-GeneSeq 17, all entries



	0	1	1	1	11	11	11	11	1	1	1	1
SCORE	0	1	3	1	6	9	12	15	18	21	24	27
STDEV	0	0	2	3	5	7	8					

PARAMETERS

Similarity matrix	Unitary	K-tuple	2
Mismatch penalty	1	Joining penalty	20
Gap penalty	1.00	Window size	27
Gap size penalty	0.05		
Cutoff score	0		
Randomization group	0		
Initial scores to save	40	Alignments to save	15
Optimized scores to save	0	Display context	100

SEARCH STATISTICS

Scores:	Mean	Median	Standard Deviation
	2	3	1.60

Times:	CPU	Total Elapsed
	00:00:23.02	00:00:26.00

Number of residues: 6065180
 Number of sequences searched: 50375
 Number of scores above cutoff: 3935

Cut-off raised to 2.
 Cut-off raised to 3.
 Cut-off raised to 4.
 Cut-off raised to 5.
 Cut-off raised to 6.

The scores below are sorted by initial score.
 Significance is calculated based on initial score.

2 100% identical sequences to the query sequence were found:

Sequence Name	Description	Length	Init. Score	Opt. Score	Sig.	Frame
1. R41975	Human T cell reactive feline	27	27	27	15.60	0
2. R36542	Peptide X.	27	27	27	15.60	0

9 100% similar sequences to the query sequence were found:

Sequence Name	Description	Length	Init. Score	Opt. Score	Sig.	Frame
3. R12120	TRFP chain 1 with leader B.	96	27	27	15.60	0
4. R27368	TRFP Chain #1 with C1 leader	96	27	27	15.60	0
5. R36548	Recombitope YZX.	96	27	27	15.60	0
6. R27367	TRFP Chain #1 with C1 leader	94	27	27	15.60	0
7. R12119	TRFP chain 1 with leader A.	94	27	27	15.60	0
8. R36539	TRFP chain 1 (with Leader A).	92	27	27	15.60	0
9. R41983	Human T cell reactive feline	92	27	27	15.60	0
10. R36540	TRFP chain 1 (with Leader B).	88	27	27	15.60	0
11. R41984	Human T cell reactive feline	88	27	27	15.60	0

The list of other best scores is:

Init. Opt.

**** 4 standard deviations above mean ****						
12. R48678	Insecticidal protoxin.	1157	10	11	4.99	0
13. R52578	Glucanase of Hordeum vulgare	334	9	9	4.37	0
14. P93413	Carbamate hydrolase.	493	9	11	4.37	0
**** 3 standard deviations above mean ****						
15. R49554	Corynebacterium halohydrin ep	244	8	10	3.74	0
16. R28296	Halohydrin epoxidase enzyme.	244	8	10	3.74	0
17. R03623	Zucchini yellow mosaic virus	283	8	10	3.74	0
18. P92062	Sequence of Isopenicillin N s	333	8	9	3.74	0
19. R45741	Myoinositol dehydrogenase.	334	8	9	3.74	0
20. R60654	pstS variant.	346	8	10	3.74	0
21. R60653	pstS variant.	346	8	10	3.74	0
22. R60652	pstS variant.	346	8	10	3.74	0
23. R60651	pstS variant.	346	8	10	3.74	0
24. R60650	pstS variant.	346	8	10	3.74	0
25. R60649	pstS variant.	346	8	10	3.74	0
26. R60648	pstS variant.	346	8	10	3.74	0
27. R60647	pstS variant.	346	8	10	3.74	0
28. R51473	pstS gene product of E.coli.	346	8	10	3.74	0
29. R60646	pstS variant.	346	8	10	3.74	0
30. R60645	pstS variant.	346	8	10	3.74	0
31. R60644	pstS variant.	346	8	10	3.74	0
32. R60643	pstS variant.	346	8	10	3.74	0
33. R60642	pstS variant.	346	8	10	3.74	0
34. R60641	pstS variant.	346	8	10	3.74	0
35. R60640	pstS variant.	346	8	10	3.74	0
36. P82053	Outer membrane protein F of P	350	8	8	3.74	0
37. R42064	Endoglucanase enzyme.	376	8	9	3.74	0
38. R37151	Dye transfer inhibiting comps	376	8	9	3.74	0
39. R27969	Endoglucanase enzyme.	376	8	9	3.74	0
40. R25429	Cellulase contained in a dete	376	8	9	3.74	0

1. US-08-300-510-1 (1-27)

R41975 Human T cell reactive feline protein fragment X.

ID R41975 standard; peptide; 27 AA.
AC R41975;
DT 21-APR-1994 (first entry)
DE Human T cell reactive feline protein fragment X.
KW Human; T cell; reactive; feline; protein; immune response; antigen;
KW tolerance; mammal; Dermatophagoides; Felis; Ambrosia; Lolium; Canis;
KW Cryptomeria; Alternaria; Alder; Betula; Quercus; Olea; Artemesia;
KW Plantago; Parietaria; Blattella; Apis; Periplaneta; autoantigen; ss.
OS Homo sapiens.
PN W09319178-A.
PD 30-SEP-1993.
PF 25-MAR-1993; U02462.
PR 25-MAR-1992; US-857311.
PR 15-MAY-1992; US-884718.
PR 15-JAN-1993; US-006116.
PA (IMMU-) IMMUNOLOGIC PHARM CORP.
PI Briner TJ, Garman RD, Gefter ML, Greenstein JL;
PI Kuo M, Morville M;
DR WPI; 93-320744/40.
PT New peptide(s) for inducing tolerance - comprise one or more
PT epitope(s) of an allergen administered subcutaneously, for
PT treating sensitivity to cats, bees, etc.
PS Claim 1; Fig 3; 107pp; English.
CC The sequences given in R41975-82 are peptides derived from a human T
CC cell reactive feline protein. These peptides are used in a
CC therapeutic composition which is useful in treating diseases which
CC involve an immune response to a protein antigen. This composition
CC may be used to induce tolerance in a mammal to Dermatophagoides,

CC Felis, Ambrosia, Lolium, Cryptomeria, Alternaria, Alder, Betula,
 CC Quercus, Olea, Artemesia, Plantago, Parietaria, Canis, Blattella,
 CC Apis, Periplaneta and to autoantigens in humans.
 SQ Sequence 27 AA;
 SQ 2 A; 1 R; 0 N; 3 D; 0 B; 0 C; 2 Q; 2 E; 0 Z; 1 G; 0 H;
 SQ 0 I; 3 L; 2 K; 0 M; 1 F; 2 P; 0 S; 2 T; 0 W; 2 Y; 4 V;

Initial Score = 27 Optimized Score = 27 Significance = 15.60
 Residue Identity = 100% Matches = 27 Mismatches = 0
 Gaps = 0 Conservative Substitutions = 0

```

X      10      20      X
KRDVDLFLTGTPEYVEQVAQYKALPV
|||||
KRDVDLFLTGTPEYVEQVAQYKALPV
X      10      20      X

```

2. US-08-300-510-1 (1-27)

R36542 Peptide X.

ID R36542 standard; Protein: 27 AA.
 AC R36542;
 DT 12-AUG-1993 (first entry)
 DE Peptide X.
 KW Human T cell reactive feline protein; TRFP; epitope; recombiteope.
 OS Felis.
 PN W09308280-A.
 PD 29-APR-1993.
 PF 16-OCT-1992; U08694.
 PR 16-OCT-1991; US-777859.
 PR 13-DEC-1991; US-807529.
 PA (IMMU-) IMMULOGIC PHARM CORP.
 PI Bond JF, Garman RD, Kuo M, Morgenstern JP, Morville M;
 PI Rogers BL;
 DR WPI; 93-152473/18.
 PT Recombiteope peptide having T-cell stimulating activity - for the
 PT diagnosis and treatment of sensitivity to protein allergens,
 PT auto:antigens and protein antigens
 PS Disclosure; Fig 4; 73pp; English.
 CC Chains 1 and 2 of the TRFP have been recombinantly expressed in E.
 CC coli and purified. T cell epitope studies using overlapping peptide
 CC regions derived from the TRFP amino acids sequence were used to
 CC identify multiple T cell epitopes in each chain of TRFP. DNA
 CC constructs were assembled in which 3 regions (encoding peptides X,
 CC Y and Z) were linked to produce DNA constructs encoding recombiteope-
 CC peptides.
 SQ Sequence 27 AA;
 SQ 2 A; 1 R; 0 N; 3 D; 0 B; 0 C; 2 Q; 2 E; 0 Z; 1 G; 0 H;
 SQ 0 I; 3 L; 2 K; 0 M; 1 F; 2 P; 0 S; 2 T; 0 W; 2 Y; 4 V;

Initial Score = 27 Optimized Score = 27 Significance = 15.60
 Residue Identity = 100% Matches = 27 Mismatches = 0
 Gaps = 0 Conservative Substitutions = 0

```

X      10      20      X
KRDVDLFLTGTPEYVEQVAQYKALPV
|||||
KRDVDLFLTGTPEYVEQVAQYKALPV
X      10      20      X

```

3. US-08-300-510-1 (1-27)

R12120 TRFP chain 1 with leader B.

ID R12120 standard; Protein: 96 AA.

NC n1c1c07
 DT 26-JUL-1991 (first entry)
 DE TRFP chain 1 with leader B.
 KW Human T cell reactive feline protein; cat allergens.
 OS Felis catus.
 FH Key Location/Qualifiers
 FT Peptide 9..26
 FT /label= Leader B
 FT Protein 27..96
 FT /label= TRFP Chain 1
 PN W09106571-A.
 PD 16-MAY-1991.
 PF 02-NOV-1990; U06548.
 PR 03-NOV-1989; US-431565.
 PA (IMMU-) IMMULOGIC PHARM COR.
 PI Geffer ML, Garman RD, Greenstein JL, Juo M, Rogers BL;
 PI Brauer AW;
 DR WPI; 91-164136/22.
 DR N-PSDB; Q11837.
 PT New pure covalently linked human T cell reactive feline protein -
 PT and modified peptide(s), used to reduce effects of cat allergens
 PT and to diagnose sensitivity to allergens.
 PS Claim 2; Fig 1; 70pp; English.
 CC Poly-A mRNA from cat parotid and mandibular glands was used to
 CC produce cDNA clones for both chain 1 and chain 2 of TRFP. These
 CC clones were then used to screen a cat genomic library. Chain 1
 CC exists in two forms having different leader sequences (A and B).
 CC The sequence can be used to express the protein and peptide derivs.
 CC which stimulate T-cells in persons allergic to cats. The peptides
 CC can be used to reduce/eliminate the allergic response partic. by
 CC modificn. of lymphokine prodn. by the T-cells. They can also be
 CC used to identify epitopes responsible for sensitivity. The DNA can
 CC be used to detect comparable sequence in other species, and also
 CC for prodn. of modified forms of TRFP esp. showing reduced binding
 CC to IgE and thus reduced tendency to cause adverse reactions.
 CC See also R12119-R12123.
 SQ Sequence 96 AA:
 SQ 12 A; 4 R; 3 N; 8 D; 0 B; 6 C; 2 Q; 7 E; 0 Z; 1 G; 0 H;
 SQ 3 I; 11 L; 7 K; 2 M; 1 F; 7 P; 3 S; 6 T; 2 W; 3 Y; 8 V;

Initial Score = 27 Optimized Score = 27 Significance = 15.60
 Residue Identity = 100% Matches = 27 Mismatches = 0
 Gaps = 0 Conservative Substitutions = 0

X 10 20 X
 KRQVDLFLTGTPDEYVEQVAQYKALPV
 |||||
 AWRCSWKRM LDAALPPCPTVAATADCEICPAVKRDVDLFLTGTPDEYVEQVAQYKALPVVLENARILKNCVD
 10 20 30 X 40 50 60 70
 AKMTEEDKENALSLLDKIYTSPLC
 80 90

4. US-08-300-510-1 (1-27)
 R27368 TRFP Chain #1 with C1 leader B sequence.

ID R27368 standard; protein; 96 AA.
 AC R27368;
 DT 25-FEB-1993 (first entry)
 DE TRFP Chain #1 with C1 leader B sequence.
 KW T cell reactive feline protein; cat allergy; allergic; IgE;
 KW desensitizing;
 OS Felis domesticus.
 FH Key Location/Qualifiers
 FT Peptide 1..27
 --

FT Protein 28..96
 FT /label= TRFP chain #1
 PN W09215613-A.
 PD 17-SEP-1992.
 PF 20-FEB-1992; U01344.
 PR 28-FEB-1991; US-662193.
 PA (IMMU-) IMMULOGIC PHARM CORP.
 PI Bond J, Kuo M;
 DR WPI; 92-331670/40.
 PT Modified human T-cell reactive feline protein - stimulates T-cell
 PT in individuals allergic to cats and shows reduced
 PT histamine-releasing properties
 PS Claim 1; Fig 1; 35pp; English.
 CC This sequence represents a modified human T-cell reactive feline
 CC protein which stimulates T-cells from an individual who is allergic
 CC to cats, but which interacts with human IgE to a lesser extent than
 CC does affinity purified TRFP. The protein is modified by treating
 CC with either a mild alkali (pH 12.5-13.5, KOH, NaOH, LiOH or tertiary
 CC amines) or an enzyme which removes O-linked groups (carbohydrate
 CC moieties). It is useful in desensitising people who are allergic to cats.
 SQ Sequence 96 AA;
 SQ 12 A; 4 R; 3 N; 8 D; 1 B; 6 C; 2 Q; 7 E; 0 Z; 1 G; 0 H;
 SQ 3 I; 11 L; 7 K; 2 M; 1 F; 7 P; 3 S; 6 T; 2 W; 3 Y; 7 V;

Initial Score = 27 Optimized Score = 27 Significance = 15.60
 Residue Identity = 100% Matches = 27 Mismatches = 0
 Gaps = 0 Conservative Substitutions = 0

X 10 20 X
 KRDVDLFLTGTPEYVEQVAQYKALPV
 ||||||||||||||||||||||||
 AWRCSWKRM LDAALPPCPTBAATADCEICPAVKR DVDLFLTGTPEYVEQVAQYKALPVVLENARILKNCVD
 10 20 30 X 40 50 60 70
 AKMTEEDKENALSLLDKIYTSPLC
 80 90

5. US-08-300-510-1 (1-27)
 R36548 Recombitope YZX.

ID R36548 standard; Protein; 96 AA.
 AC R36548;
 DT 12-AUG-1993 (first entry)
 DE Recombitope YZX.
 KW Human T cell reactive feline protein; TRFP; epitope; recombiteope
 KW sensitivity; Felis domesticus.
 OS Synthetic.
 FH Key Location/Qualifiers
 FT Cleavage_site 14..15
 FT /label= thrombin_cleavage_site
 PN W09308280-A.
 PD 29-APR-1993.
 PF 16-OCT-1992; U08694.
 PR 16-OCT-1991; US-777859.
 PR 13-DEC-1991; US-807529.
 PA (IMMU-) IMMULOGIC PHARM CORP.
 PI Bond JF, Garman RD, Kuo M, Morgenstern JP, Morville M;
 PI Rogers BL;
 DR WPI; 93-152473/18.
 DR N-PSDB; 041572.
 PT Recombitope peptide having T-cell stimulating activity - for the
 PT diagnosis and treatment of sensitivity to protein allergens,
 PT auto:antigens and protein antigens
 PS Disclosure; Fig 8; 73pp; English.

CC Preferred recombinant peptides for treating sensitivity to Felis
 CC domesticus are derived from the the genus Felis and comprise
 CC regions selected from peptides X, Y, Z, A and B, of TRFP, and
 CC modifications thereof, such as peptide C.
 CC Oligonucleotides C, D, E, F, G, H and I are used in the
 CC construction of recombinant peptide YZX.
 SQ Sequence 96 AA:
 SQ 8 A; 4 R; 5 N; 6 D; 0 B; 1 C; 2 Q; 10 E; 0 Z; 4 G; 6 H;
 SQ 1 I; 12 L; 7 K; 2 M; 4 F; 5 P; 2 S; 5 T; 0 W; 2 Y; 10 V;

Initial Score = 27 Optimized Score = 27 Significance = 15.60
 Residue Identity = 100% Matches = 27 Mismatches = 0
 Gaps = 0 Conservative Substitutions = 0

X
KRD
|||

MGHHHHHHEFLVPRGSKALPVVLENARILKNCVDAKMTTEDEKFFAVANGNELLLDLSLTKVNATEPERKRD
 10 20 30 40 50 60 70

10 20 X
 VDLFLTGTPEYVEQVAQYKALPV
 |||||
 VDLFLTGTPEYVEQVAQYKALPV
 80 90 X

6. US-08-300-510-1 (1-27)

R27367 TRFP Chain #1 with C1 leader A sequence.

ID R27367 standard; protein; 94 AA.
 AC R27367;
 DT 25-FEB-1993 (first entry)
 DE TRFP Chain #1 with C1 leader A sequence.
 KW T cell reactive feline protein.
 OS Felis domesticus.
 FH Key Location/Qualifiers
 FT Peptide 1..25
 FT /label= Leader A
 FT Protein 25..94
 FT /label= TRFP chain #1
 PN W09215613-A.
 PD 17-SEP-1992.
 PF 20-FEB-1992; U01344.
 PR 28-FEB-1991; US-662193.
 PA (IMMU-) IMMULOGIC PHARM CORP.
 PI Bond J, Kuo M;
 DR WPI; 92-331670/40.
 PT Modified human T-cell reactive feline protein - stimulates T-cell
 PT in individuals allergic to cats and shows reduced
 PT histamine-releasing properties
 PS Claim 1; Fig 1; 35pp; English.
 CC This sequence represents a modified human T-cell reactive feline
 CC protein which stimulates T-cells from an individual who is allergic
 CC to cats, but which interacts with human IgE to a lesser extent than
 CC does affinity purified TRFP. The protein is modified by treating
 CC with either a mild alkali (pH 12.5-13.5, KOH, NaOH, LiOH or tertiary
 CC amines) or an enzyme which removes O-linked groups (carbohydrate
 CC moieties). It is useful in desensitising people who are allergic to cats.
 SQ Sequence 94 AA:
 SQ 9 A; 3 R; 4 N; 6 D; 0 B; 5 C; 2 Q; 7 E; 0 Z; 4 G; 0 H;
 SQ 5 I; 15 L; 7 K; 2 M; 1 F; 4 P; 2 S; 4 T; 2 W; 3 Y; 9 V;

Initial Score = 27 Optimized Score = 27 Significance = 15.60
 Residue Identity = 100% Matches = 27 Mismatches = 0
 Gaps = 0 Conservative Substitutions = 0

```

      X      10      20      X
      KRVDLFLTGTTPDEYVEQVAQYKALPV
      |||||
CIMKGARVLVLLWAALLLIWGGNCEICPAVKRDVDLFLTGTTPDEYVEQVAQYKALPVVLENARILKNCVDAK
      10      20      30      40      50      X 60      70

MTEEDKENALSLLDKIYTSPLC
      80      90

```

7. US-08-300-510-1 (1-27)

R12119 TRFP chain 1 with leader A.

ID R12119 standard; Protein: 94 AA.
AC R12119;
DT 26-JUL-1991 (first entry)
DE TRFP chain 1 with leader A.
KW Human T cell reactive feline protein; cat allergens.
OS Felis catus.
FH Key Location/Qualifiers
FT Peptide 3..24
FT /label= Leader B
FT Protein 25..94
FT /label= TRFP Chain 1
PN W09106571-A.
PD 16-MAY-1991.
PF 02-NOV-1990; U06548.
PR 03-NOV-1989; US-431565.
PA (IMMU-) IMMULOGIC PHARM COR.
PI Gefter ML, Garman RD, Greenstein JL, Juo M, Rogers BL;
PI Brauer AW;
DR WPI; 91-164136/22.
DR N-PSDB; Q11836.
PT New pure covalently linked human T cell reactive feline protein -
PT and modified peptide(s), used to reduce effects of cat allergens
PT and to diagnose sensitivity to allergens.
PS Claim 2; Fig 1; 70pp; English.
CC Poly-A mRNA from cat parotid and mandibular glands was used to
CC produce cDNA clones for both chain 1 and chain 2 of TRFP. These
CC clones were then used to screen a cat genomic library. Chain 1
CC exists in two forms having different leader sequences (A and B).
CC The sequence can be used to express the protein and peptide derivs.
CC which stimulate T-cells in persons allergic to cats. The peptides
CC can be used to reduce/eliminate the allergic response partic. by
CC modifcn. of lymphokine prodn. by the T-cells. They can also be
CC used to identify epitopes responsible for sensitivity. The DNA can
CC be used to detect comparable sequence in other species, and also
CC for prodn. of modified forms of TRFP esp. showing reduced binding
CC to IgE and thus reduced tendency to cause adverse reactions.
CC See also R12120-R12123.
SQ Sequence 94 AA;
SQ 9 A; 3 R; 4 N; 6 D; 0 B; 5 C; 2 Q; 7 E; 0 Z; 4 G; 0 H;
SQ 5 I; 15 L; 7 K; 2 M; 1 F; 4 P; 2 S; 4 T; 2 W; 3 Y; 9 V;

Initial Score	=	27	Optimized Score	=	27	Significance	=	15.60
Residue Identity	=	100%	Matches	=	27	Mismatches	=	0
Gaps	=	0	Conservative Substitutions	=			=	0

```

      X      10      20      X
      KRVDLFLTGTTPDEYVEQVAQYKALPV
      |||||
CIMKGARVLVLLWAALLLIWGGNCEICPAVKRDVDLFLTGTTPDEYVEQVAQYKALPVVLENARILKNCVDAK
      10      20      30      40      50      X 60      70

MTEEDKENALSLLDKIYTSPLC

```


8. US-08-300-510-1 (1-27)

R36539 TRFP chain 1 (with Leader A).

ID R36539 standard; Protein; 92 AA.
 AC R36539;
 DT 12-AUG-1993 (first entry)
 DE TRFP chain 1 (with Leader A).
 KW Human T cell reactive feline protein; TRFP; leader A; leader B;
 KW epitope.
 OS Felis.
 FH Key Location/Qualifiers
 FT Peptide 1..22
 FT /label= leader_peptide
 PN W09308280-A.
 PD 29-APR-1993.
 PF 16-OCT-1992; U08694.
 PR 16-OCT-1991; US-777859.
 PR 13-DEC-1991; US-807529.
 PA (IMMU-) IMMULOGIC PHARM CORP.
 PI Bond JF, Garman RD, Kuo M, Morgenstern JP, Morville M;
 PI Rogers BL;
 DR WPI; 93-152473/18.
 DR N-PSDB; Q41556.
 PT Recombitope peptide having T-cell stimulating activity - for the
 PT diagnosis and treatment of sensitivity to protein allergens,
 PT auto:antigens and protein antigens
 PS Disclosure; Fig 1; 73pp; English.
 CC Chains 1 and 2 of the TRFP have been recombinantly expressed in E.
 CC coli and purified. T cell epitope studies using overlapping peptide
 CC regions derived from the TRFP amino acids sequence were used to
 CC identify multiple T cell epitopes in each chain of TRFP.
 SQ Sequence 92 AA;
 SQ 9 A; 3 R; 4 N; 6 D; 0 B; 4 C; 2 Q; 7 E; 0 Z; 4 G; 0 H;
 SQ 4 I; 15 L; 7 K; 2 M; 1 F; 4 P; 2 S; 4 T; 2 W; 3 Y; 9 V;

Initial Score = 27 Optimized Score = 27 Significance = 15.60
 Residue Identity = 100% Matches = 27 Mismatches = 0
 Gaps = 0 Conservative Substitutions = 0

X 10 20 X
 KRVDLFLTGTPDEYVEGVAQYKALPV
 ||||||||||||||||||||||||||||
 MKGARVLVLLWAAALLLINGGNCEICPAVKRDVDLFLTGTPDEYVEGVAQYKALPVLLENARILKNCVDAKMT
 10 20 30 40 50 X 60 70
 EEDKENALSLLDKIYTSPLC
 80 90

9. US-08-300-510-1 (1-27)

R41983 Human T cell reactive feline protein A chain 1.

ID R41983 standard; Protein; 92 AA.
 AC R41983;
 DT 21-APR-1994 (first entry)
 DE Human T cell reactive feline protein A chain 1.
 KW Human; T cell; reactive; feline; protein; immune response; antigen;
 KW tolerance; mammal; Dermatophagoides; Felis; Ambrosia; Lolium; Canis;
 KW Cryptomeria; Alternaria; Alder; Betula; Quercus; Olea; Artemesia;
 KW Plantago; Parietaria; Blattella; Apis; Periplaneta; autoantigen.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT Peptide 1..22

FT /note= "signal peptide"
 FT Protein 23..92
 FT /note= "Mature protein"
 PN W09319178-A.
 PD 30-SEP-1993.
 PF 25-MAR-1993; U02462.
 PR 25-MAR-1992; US-857311.
 PR 15-MAY-1992; US-884718.
 PR 15-JAN-1993; US-006116.
 PA (IMMU-) IMMUNOLOGIC PHARM CORP.
 PI Briner TJ, Garman RD, Gefter ML, Greenstein JL, Kuo M;
 PI Morville M;
 DR WPI; 93-320744/40.
 DR N-PSDB; 049533.
 PT New peptide(s) for inducing tolerance - comprise one or more
 PT epitope(s) of an allergen administered subcutaneously, for
 PT treating sensitivity to cats, bees, etc.
 PS Disclosure; Fig 1; 107pp; English.
 CC The sequences given in R41983-84 represent chain 1 of human T cell
 CC reactive feline proteins (TRFP) A and B respectively. Peptides
 CC derived from TRFP may be used in a therapeutic composition which is
 CC useful in treating diseases which involve an immune response to a
 CC protein antigen. This composition may be used to induce tolerance
 CC in a mammal to Dermatophagoides, Felis, Ambrosia, Lolium, Cryptomeria,
 CC Alternaria, Alder, Betula, Quercus, Olea, Artemesia, Plantago,
 CC Parietaria, Canis, Blattella, Apis, Periplaneta and to autoantigens
 CC in humans.
 SQ Sequence 92 AA;
 SQ 9 A; 3 R; 4 N; 6 D; 0 B; 4 C; 2 Q; 7 E; 0 Z; 4 G; 0 H;
 SQ 4 I; 15 L; 7 K; 2 M; 1 F; 4 P; 2 S; 4 T; 2 W; 3 Y; 9 V;

Initial Score = 27 Optimized Score = 27 Significance = 15.60
 Residue Identity = 100% Matches = 27 Mismatches = 0
 Gaps = 0 Conservative Substitutions = 0

X 10 20 X
 KRQVDLFLTGTPDEYVEQVAQYKALPV
 |||||
 MKGARVLVLLWAALLLIWGGNCEICPAVKRQVDLFLTGTPDEYVEQVAQYKALPVVLENARILKNCVDAKMT
 10 20 30 40 50 X 60 70
 EEDKENALSLLDKIYTSPLC
 80 90

10. US-08-300-510-1 (1-27)
 R36540 TRFP chain 1 (with Leader B).

ID R36540 standard; Protein; 88 AA.
 AC R36540;
 DT 12-AUG-1993 (first entry)
 DE TRFP chain 1 (with Leader B).
 KW Human T cell reactive feline protein; TRFP; leader A; leader B;
 KW epitope.
 OS Felis.
 FH Key Location/Qualifiers
 FT Peptide 1..18
 FT /label= leader_peptide
 PN W09308280-A.
 PD 29-APR-1993.
 PF 16-OCT-1992; U08694.
 PR 16-OCT-1991; US-777859.
 PR 13-DEC-1991; US-807529.
 PA (IMMU-) IMMUNOLOGIC PHARM CORP.
 PI Bond JF, Garman RD, Kuo M, Morgenstern JP, Morville M;
 PI Rogers BL;

UN WPI: 93-1524/3/18.
 DR N-PSDB: Q41557.
 PT Recombitope peptide having T-cell stimulating activity - for the
 PT diagnosis and treatment of sensitivity to protein allergens,
 PT auto:antigens and protein antigens
 PS Disclosure: Fig 1; 73pp; English.
 CC Chains 1 and 2 of the TRFP have been recombinantly expressed in E.
 CC coli and purified. T cell epitope studies using overlapping peptide
 CC regions derived from the TRFP amino acids sequence were used to
 CC identify multiple T cell epitopes in each chain of TRFP.
 SQ Sequence 88 AA:
 SQ 11 A; 2 R; 3 N; 8 D; 0 B; 5 C; 2 Q; 7 E; 0 Z; 1 G; 0 H;
 SQ 3 I; 11 L; 6 K; 2 M; 1 F; 7 P; 2 S; 6 T; 0 W; 3 Y; 8 V;

Initial Score = 27 Optimized Score = 27 Significance = 15.60
 Residue Identity = 100% Matches = 27 Mismatches = 0
 Gaps = 0 Conservative Substitutions = 0

X 10 20 X
 KRQVDLFLTGTPDEYVEQVAQYKALPV
 |||||
 MLDAALPPCPTVAATADCEICPAVKRDVDLFLTGTPDEYVEQVAQYKALPVVLENARILKNCVDAKMTEEDK
 10 20 X 30 40 50 60 70
 ENALSLLDKIYTSPLC
 80

11. US-08-300-510-1 (1-27)
 R41984 Human T cell reactive feline protein B chain 1.

ID R41984 standard; Protein; 88 AA.
 AC R41984;
 DT 21-APR-1994 (first entry)
 DE Human T cell reactive feline protein B chain 1.
 KW Human; T cell; reactive; feline; protein; immune response; antigen;
 KW tolerance; mammal; Dermatophagoides; Felis; Ambrosia; Lolium; Canis;
 KW Cryptomeria; Alternaria; Alder; Betula; Quercus; Olea; Artemisia;
 KW Plantago; Parietaria; Blattella; Apis; Periplaneta; autoantigen.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT Peptide 1..17
 FT /note= "Signal peptide"
 FT Protein 18..88
 FT /note= "Mature protein"
 PN W09319178-A.
 PD 30-SEP-1993.
 PF 25-MAR-1993; U02462.
 PR 25-MAR-1992; US-857311.
 PR 15-MAY-1992; US-884718.
 PR 15-JAN-1993; US-006116.
 PA (IMMU-) IMMUNOLOGIC PHARM CORP.
 PI Briner TJ, Garman RD, Gefter ML, Greenstein JL, Kuo M;
 PI Morville M;
 DR WPI: 93-320744/40.
 DR N-PSDB: Q49534.
 PT New peptide(s) for inducing tolerance - comprise one or more
 PT epitope(s) of an allergen administered subcutaneously, for
 PT treating sensitivity to cats, bees, etc.
 PS Disclosure: Fig 1; 107pp; English.
 CC The sequences given in R41983-84 represent chain 1 of human T cell
 CC reactive feline proteins (TRFP) A and B respectively. Peptides
 CC derived from TRFP may be used in a therapeutic composition which is
 CC useful in treating diseases which involve an immune response to a
 CC protein antigen. This composition may be used to induce tolerance
 CC in a mammal to Dermatophagoides, Felis, Ambrosia, Lolium, Cryptomeria,

CC Alternaria, Alder, Betula, Quercus, Ulex, Artemisia, Plantago,
CC Parietaria, Canis, Blattella, Apis, Periplaneta and to autoantigens
CC in humans.

SQ Sequence 88 AA:
SQ 11 A; 2 R; 3 N; 8 D; 0 B; 5 C; 2 Q; 7 E; 0 Z; 1 G; 0 H;
SQ 3 I; 11 L; 6 K; 2 M; 1 F; 7 P; 2 S; 6 T; 0 W; 3 Y; 8 V;

Initial Score = 27 Optimized Score = 27 Significance = 15.60
Residue Identity = 100% Matches = 27 Mismatches = 0
Gaps = 0 Conservative Substitutions = 0

X 10 20 X
KRDVDLFLTGTPEYVEQVAQYKALPV
|||||
MLDAALPPCPTVAATADCEICPAVKRDVDLFLTGTPEYVEQVAQYKALPVVLENARILKNCVDAKMTEEDK
10 20 X 30 40 50 60 70

ENALSLLDKIYTSPLC
80

12. US-08-300-510-1 (1-27)
R48678 Insecticidal protoxin.

ID R48678 standard; Protein; 1157 AA.
AC R48678;
DT 13-OCT-1994 (first entry)
DE Insecticidal protoxin.
KW Insecticide; toxin; protoxin; Lepidoptera; pest control;
KW Bacillus thuringiensis; crop protection; crystal protein;
KW delta endotoxin.
OS Bacillus thuringiensis (BTS02618A).
PN W09405771-A.
PD 17-MAR-1994.
PF 12-JUL-1993; E01820.
PR 27-AUG-1992; EP-402358.
PR 09-APR-1993; EP-400949.
PA (PLBZ) PLANT GENETIC SYSTEMS NV.
PI Jansens S, Lambert B, Peferoen M, Van Audenhove K;
DR WPI; 94-101176/12.
DR N-PSDB; Q56782.
PT Bacillus thuringiensis strains producing insecticidal proteins
PT active against Lepidoptera species - to produce transgenic
PT plants resistant to Lepidoptera species
PS Claim 3; Page 39-44; 49pp; English.
CC The DNA encoding the protoxin can be used to transform a plant to
CC protect the plant from Lepidopteran pests. The protoxin produced
CC yields a toxin product after trypsin digestion. The protoxin, toxin,
CC crystal proteins and the Bacillus strain producing them can all be
CC used as the active ingredient in an insecticidal composition.
SQ Sequence 1157 AA:
SQ 74 A; 71 R; 82 N; 75 D; 0 B; 12 C; 59 Q; 66 E; 0 Z; 80 G; 22 H;
SQ 50 I; 106 L; 24 K; 13 M; 48 F; 46 P; 83 S; 89 T; 15 W; 54 Y; 88 V;

Initial Score = 10 Optimized Score = 11 Significance = 4.99
Residue Identity = 46% Matches = 13 Mismatches = 13
Gaps = 2 Conservative Substitutions = 0

NPGVDGNTNRIESTAVDFRSALIGIYGVNRSFVPGGLFNGTTSPANGGCRDLYDTNDELPPDESTGSSTHRL
410 420 430 440 450 460 470
X 10 20 X
KRDVDLFLTGTPEYVEQVAQYKA-LPV
||||| ||| | |||
SHVTFSSFQTNQAGSIANAGSVPTYVWTRRDVDLNNITP-NRITQLPLVKASAPVSGTTVLKCPGFTGGGI
480 490 500 510 520 530 540

LRRITNGTGTGLRVTVNSPLTQQYRLRVRFASGTGNFSIRVLRGGVSIQDVRLGSTMNRGQELTYESFFTREF
 550 560 570 580 590 600 610

TTTGPFNPPFTF
 620

13. US-08-300-510-1 (1-27)
 R52578 Glucanase of Hordeum vulgare L.

ID R52578 standard; Protein; 334 AA.
 AC R52578;
 DT 05-DEC-1994 (first entry)
 DE Glucanase of Hordeum vulgare L.
 KW Antifungal; pathogen; resistance; transgenic organism; synergy;
 KW crop protection; transgenic plant; chitinase; glucanase;
 KW protein synthesis inhibitor; disease.
 OS Hordeum vulgare L.
 PN DE4234131-A.
 PD 21-APR-1994.
 PF 09-OCT-1992; 234131.
 PR 09-OCT-1992; DE-234131.
 PA (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
 PI Chet I, Eckes P, Gornhardt B, Jach G, Logemann J;
 PI Mundy J, Schellj, Goernhardt B;
 DR WPI; 94-136599/17.
 DR N-PSDB; R52578.
 PT Transgenic organisms contg. at least 2 pathogen inhibiting genes
 PT - esp. plants contg. genes with antifungal activity, show
 PT synergistic increase in disease resistance, also new DNA transfer
 PT vectors
 PS Example 2; Page 15-16; 19pp; German.
 CC Glucanase is an enzyme which breaks down glucan, a glucose polymer
 CC present in fungal cell walls. The sequence encoding the glucanase
 CC enzyme may be used in the construction of transgenic organisms,
 CC especially plants, to produce pathogen resistant organisms. The
 CC genome of such transgenic organisms preferably contains more than
 CC one gene with pathogen inhibiting activity, each gene under the
 CC control of active promoters. The two gene products then show a
 CC synergistic increase in pathogen induced activity so that the
 CC transgenic organisms have a greater degree of resistance or
 CC resistance against a wider spectrum of diseases.
 SQ Sequence 334 AA;
 SQ 50 A; 14 R; 28 N; 13 D; 0 B; 1 C; 12 Q; 8 E; 0 Z; 31 G; 1 H;
 SQ 19 I; 22 L; 10 K; 6 M; 16 F; 16 P; 27 S; 16 T; 2 W; 16 Y; 26 V;

Initial Score = 9 Optimized Score = 9 Significance = 4.37
 Residue Identity = 33% Matches = 9 Mismatches = 18
 Gaps = 0 Conservative Substitutions = 0

NEVQGGATQSILPAMRNLSAAGLGAIKVSTSIKFEVANSFPSPAGVFNNAYITDVARLLASTGAPLL
 130 140 150 160 170 180 190
 X 10 20 X
 KRVDLFLTGTPEYVEQVAQYKALPV
 || || | || ||
 ANVYPYFAYRDNPGSISLNYATFQPGTTVRDQNNGLTYTSLFDAMVDAVYAALEKAGAPAVKVVVSESGWPS
 200 210 220 230 240 250 260
 AGGFAASAGNARTYNQGLINHVGGGTPKKREALETYIFAMFNENQKTGDATERSFGLFNPDKSPAYNIQF
 270 280 290 300 310 320 330

14. US-08-300-510-1 (1-27)
 P93413 Carbanate hydrolase.

ID P93413 standard; protein; 493 AA.
AC P93413;
DT 27-APR-1990 (first entry)
DE Carbamate hydrolase.
KW Carbamate hydrolase; Arthrobacter oxidans; phenmedipham;
KW methyl 3-hydroxyphenylcarbamate.
OS Arthrobacter oxidans DSM 4044.
PN EP-343100-A.
PD 23-NOV-1989.
PF 17-May-1989; 730123.
PR 19-MAY-1988; DE-381738.
PA (SCHD) Schering AG.
PI Pohlenz HD, Boidol W;
DR WPI; 89-341858/47.
DR N-PSDB; N92585.
PT Pure carbamate hydrolase isolation from Arthrobacter oxidans - able to
PT destroy herbicide phenmedipham, and DNA encoding it, for imparting
PT resistance to plants.
PS Disclosure; Fig. 7; 17pp; german.
CC Purified carbamate hydrolase can be used to isolate/identify A.oxidans
CC carbamate hydrolase gene system. This system makes plants resistant to
CC the herbicide phenmedipham. Carbamate hydrolase has pH optimum 6.8, mol.
CC wt. 50-60kD, isoelectric point 6.2, and can cleave phenmedipham to methyl
CC 3-hydroxyphenylcarbamate, m-toluidine and CO2, so inactivating
CC it. It is produced by A.oxidans DSM 4044 which contains the 41 kb plasmid
CC PHP52.
SQ Sequence 493 AA;
SQ 51 A; 33 R; 11 N; 41 D; 0 B; 3 C; 15 Q; 24 E; 0 Z; 50 G; 19 H;
SQ 24 I; 40 L; 4 K; 4 M; 21 F; 39 P; 20 S; 30 T; 14 W; 10 Y; 40 V;

Initial Score = 9 Optimized Score = 11 Significance = 4.37
Residue Identity = 32% Matches = 13 Mismatches = 14
Gaps = 13 Conservative Substitutions = 0

VPYAEPPVGDLRWRAARPHAGWTGVRDASAYGPSAPQPVPEGGSPILGTHGDPPFDEDCLTLNLWTPNLDGG
30 40 50 60 70 80 90
X 10 20 X
KRDVDL-----FLTGTPDEYVEQVAGYKALPV
|| || || || || || || ||
SRPVLVWIHGGGLLTGSGNLPNYATDTFARDGDLVGISINYRLGPLGFLAGMGDENVWLTQVEALRWIADN
100 110 120 130 140 150 160 170
VAAFGGDPNRLTLVGQSGGAYSIAALAGHPVARQLFHRAILQSPFQMGPHTEESTARTKALARHLGHDDI
180 190 200 210 220 230 240
EALRHEPWERLIQGTIGVLMENHK
250 260

15. US-08-300-510-1 (1-27)
R49554 Corynebacterium halohydrin epoxidase encoded by pl

ID R49554 standard; Protein; 244 AA.
AC R49554;
DT 07-JUL-1994 (first entry)
DE Corynebacterium halohydrin epoxidase encoded by plasmid pST015.
KW 3-hydroxynitrile; halohydrin epoxidase gene; recombinant plasmid;
KW Escherichia coli.
OS Corynebacterium sp. (strain N-1074).
PN J05317066-A.
PD 03-DEC-1993.
PF 04-MAR-1991; 062597.
PR 04-MAR-1991; JP-062597.
PA (NITT) NITTO CHEM IND CO LTD.

RA YAMA / YAMADA M.
 DR WPI; 94-011029/02.
 DR N-PSDB; Q54372.
 PT Prepn. of a 3-hydroxy:nitrile cpd. - by transforming microbe with
 PT recombinant plasmid having halohydrin epoxidase gene
 PS Claim 2; Page 5; 23pp; Japanese.
 CC Microorganisms transformed by a recombinant plasmid comprising a
 CC halohydrin epoxidase gene are used for production of 3-
 CC hydroxynitrile cpds. from 1,2-epoxy cpds. The transformed E.coli
 CC JM109/pST015 encodes a protein having the amino acid sequence
 CC R49554.
 SQ Sequence 244 AA;
 SQ 34 A; 18 R; 9 N; 11 D; 0 B; 1 C; 6 Q; 16 E; 0 Z; 17 G; 4 H;
 SQ 15 I; 23 L; 5 K; 3 M; 13 F; 19 P; 12 S; 15 T; 1 W; 7 Y; 15 V;

Initial Score = 8 Optimized Score = 10 Significance = 3.74
 Residue Identity = 36% Matches = 11 Mismatches = 16
 Gaps = 3 Conservative Substitutions = 0

VFEALSIPIILLQSAIAPLRAAGGASVIFITSSVGKKPLAYNPLYGPARAATVALVESAAKTLSDGILLY
 110 120 130 140 150 160 170

X 10 20 X
 KRQVDLFLTGTPDE---YVEQVAQYKALPV

AIGPNFFNNPTYFPTSDWENNPELRERVERDVPLGRLGRPDEMGAITFLASRRRAAPIVGQFFAFTGGVLP
 180 190 200 X 210 220 230 240

> O <
 O| |O IntelliGenetics
 > O <

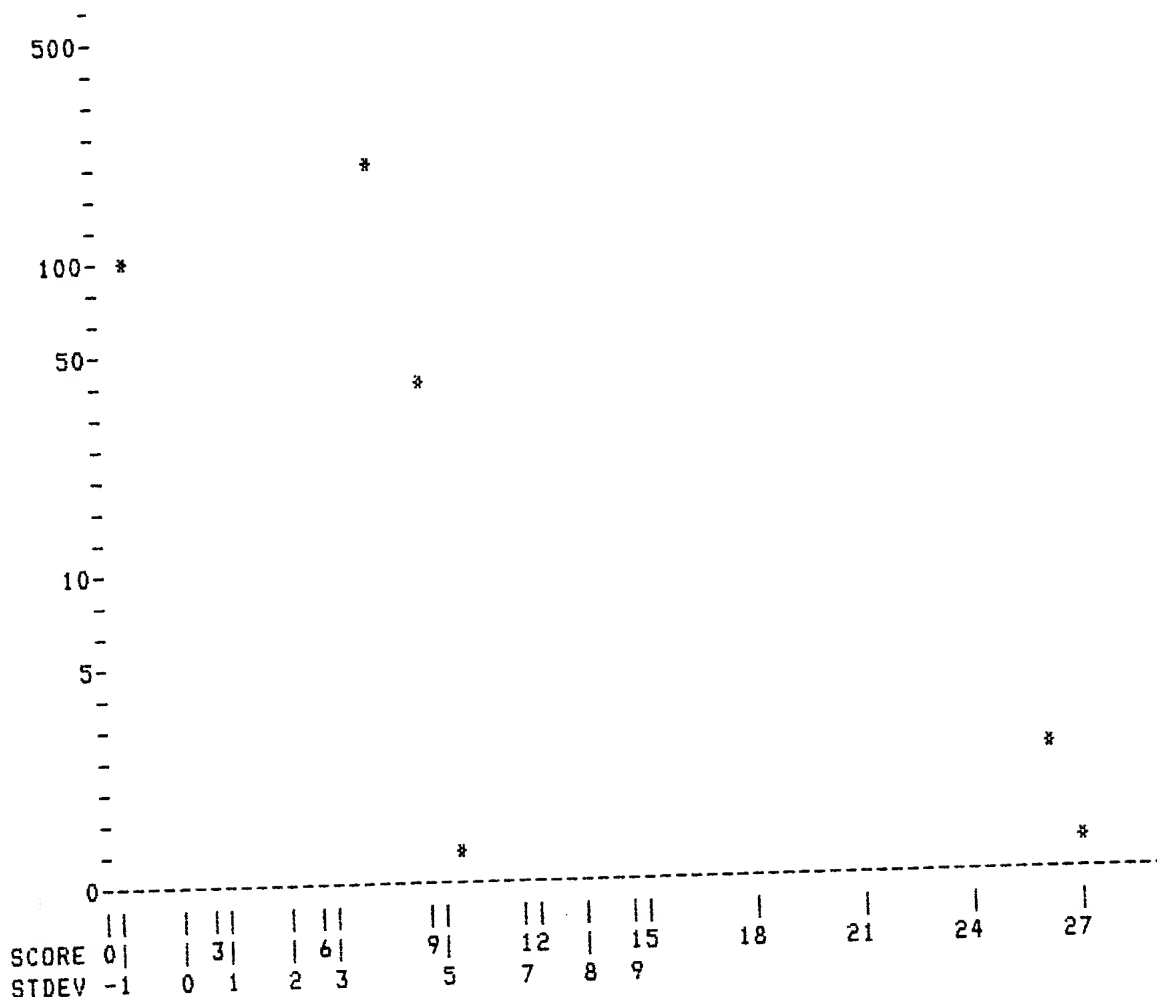
FastDB - Fast Pairwise Comparison of Sequences
 Release 5.4

Results file 1-pir.res made by on Fri 24 Mar 95 7:46:05-PST.

Query sequence being compared: US-08-300-510-1 (1-27)
 Number of sequences searched: 75511
 Number of scores above cutoff: 4166

Results of the initial comparison of US-08-300-510-1 (1-27) with:
 Data bank : PIR-43, all entries

100000-
 N -
 U50000-
 M -
 B -
 E - *
 R - * *
 D -
 F10000- *
 S -
 E 5000-
 Q - *
 U -
 E *
 N -
 C -
 E -
 S 1000-
 -



PARAMETERS

Similarity matrix	Unitary	K-tuple	2
Mismatch penalty	1	Joining penalty	20
Gap penalty	1.00	Window size	27
Gap size penalty	0.05		
Cutoff score	0		
Randomization group	0		
Initial scores to save	40	Alignments to save	15
Optimized scores to save	0	Display context	100

SEARCH STATISTICS

Scores:	Mean	Median	Standard Deviation
	3	4	1.38

Times:	CPU	Total Elapsed
	00:01:01.07	00:01:03.00

Number of residues:	22468834
Number of sequences searched:	75511
Number of scores above cutoff:	4166

Cut-off raised to 3.
 Cut-off raised to 4.
 Cut-off raised to 5.
 Cut-off raised to 6.
 Cut-off raised to 7.

The scores below are sorted by initial score.
 Significance is calculated based on initial score.

A 100% similar sequence to the query sequence was found:

Sequence Name	Description	Length	Init. Opt.		Sig.	Frame
			Score	Score		
1. A53283	major cat allergen Fel d I al	40	27	27	17.44	0

The list of other best scores is:

Sequence Name	Description	Length	Init. Score	Opt. Score	Sig.	Frame
**** 16 standard deviations above mean ****						
2. JC1126	major allergen chain 1 precur	88	26	26	16.71	0
3. JC1136	major allergen chain 1 precur	92	26	26	16.71	0
**** 5 standard deviations above mean ****						
4. ISUTTB	triose-phosphate isomerase (E	250	10	10	5.09	0
**** 4 standard deviations above mean ****						
5. PN0644	hypothetical protein 66 - Str	66	9	9	4.36	0
6. LNPG1	pulmonary surfactant protein	79	9	9	4.36	0
7. S15199	hydrogenase isozyme hypC - Es	90	9	12	4.36	0
8. F53275	kappa 1 b95 allotype=constant	104	9	11	4.36	0
9. K5RBV	Ig kappa chain C region (B5 v	104	9	11	4.36	0
10. A20968	Ig kappa-1b5 chain C region -	105	9	11	4.36	0
11. S43188	orotidine-5'-phosphate decarb	232	9	9	4.36	0
12. JS0618	glutathione transferase (EC 2	244	9	10	4.36	0
13. A38233	triose-phosphate isomerase (E	253	9	9	4.36	0
14. S04405	hydroxyneurosporene synthase	281	9	10	4.36	0
15. S21394	transposase - Mycobacterium t	308	9	10	4.36	0
16. S37652	FVT1 protein - human	332	9	9	4.36	0
17. D38664	glucan endo-1,3-beta-D-glucos	334	9	9	4.36	0
18. S05510	glucan endo-1,3-beta-D-glucos	334	9	9	4.36	0
19. A35630	regulatory protein algR3 - Ps	340	9	9	4.36	0
20. S34494	ccsA protein - Euglena gracil	348	9	9	4.36	0
21. S23088	ccsA protein - Euglena gracil	348	9	9	4.36	0
22. JQ0148	hypothetical 34.4K protein -	351	9	9	4.36	0
23. A36128	regulatory protein algP - Pse	352	9	9	4.36	0
24. S13822	protein Z4 - barley	399	9	9	4.36	0
25. S12785	protein ch-42 precursor, chlo	424	9	9	4.36	0
26. A37807	3-phosphoshikimate 1-carboxyv	450	9	9	4.36	0
27. A48788	leucyl aminopeptidase (EC 3.4	469	9	9	4.36	0
28. A45737	phenylcarbamate hydrolase - A	493	9	11	4.36	0
29. PQ0470	probable leucyl aminopeptidas	554	9	10	4.36	0
30. S41376	leucine aminopeptidase - pota	573	9	10	4.36	0
31. S22967	polyphenol oxidase precursor	604	9	9	4.36	0
32. S22965	polyphenol oxidase precursor	630	9	9	4.36	0
33. S18737	gag polyprotein - simian foam	647	9	10	4.36	0
34. S32899	ferric-pseudobactin receptor	809	9	9	4.36	0
35. S10639	fruB protein - Rhodobacter ca	827	9	10	4.36	0
36. B27211	virA protein - Agrobacterium	829	9	9	4.36	0
37. S04035	virA protein - Agrobacterium	829	9	9	4.36	0
38. WMBE56	infected cell protein ICP18.5	850	9	10	4.36	0
39. S44250	integrin alpha 5 chain - mous	1053	9	9	4.36	0
40. S35548	DNA-directed RNA polymerase (1210	9	10	4.36	0

1. US-08-300-510-1 (1-27)

A53283 major cat allergen Fel d I alpha chain - cat (frag

ENTRY A53283 #type fragment
 TITLE major cat allergen Fel d I alpha chain - cat (fragment)
 ORGANISM #formal_name Felis silvestris catus #common_name domestic cat
 DATE 12-May-1994 #sequence_revision 12-May-1994 #text_change
 12-May-1994

ACCESSION A53283
 REFERENCE A53283
 #authors Duffort, D.A.; Carreira, J.; Nitti, G.; Polo, F.; Lombardero, M.
 #journal Mol. Immunol. (1991) 28:301-309
 #title Studies on the biochemical structure of the major cat allergen *Felis domesticus* I.
 #accession A53283
 ##status preliminary
 ##molecule_type protein
 ##residues 1-40 ##label DUF
 SUMMARY #length 40 #checksum 3032
 SEQUENCE

Initial Score = 27 Optimized Score = 27 Significance = 17.44
 Residue Identity = 100% Matches = 27 Mismatches = 0
 Gaps = 0 Conservative Substitutions = 0

```

      X      10      20      X
      KRVDLFLTGTPDEYVEQVAQYKALPV
      ||||||||||||||||||||
EICPAVKRDVDLFLTGTPDEYVEQVAQYKALPVVLENARI
      X 10      20      30  X      40
  
```

2. US-08-300-510-1 (1-27)
 JC1126 major allergen chain 1 precursor B - cat

ENTRY JC1126 #type complete
 TITLE major allergen chain 1 precursor B - cat
 ORGANISM #formal_name *Felis silvestris catus* #common_name domestic cat
 DATE 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change
 31-Dec-1993
 ACCESSIONS JC1126
 REFERENCE JC1126
 #authors Griffith, I.J.; Craig, S.; Pollock, J.; Yu, X.B.;
 Morgenstern, J.P.; Rogers, B.L.
 #journal Gene (1992) 113:263-268
 #title Expression and genomic structure of the genes encoding FdI,
 the major allergen from the domestic cat.
 #accession JC1126
 ##molecule_type DNA
 ##residues 1-88 ##label GRI
 GENETICS
 #gene Ch1
 #introns 17/1; 79/3
 FEATURE
 1-18 #domain signal sequence #status predicted #label SIG\
 19-88 #product major allergen chain 1 #status predicted #label
 MAT
 SUMMARY #length 88 #molecular-weight 9586 #checksum 4095
 SEQUENCE

Initial Score = 26 Optimized Score = 26 Significance = 16.71
 Residue Identity = 96% Matches = 26 Mismatches = 1
 Gaps = 0 Conservative Substitutions = 0

```

      X      10      20      X
      KRVDLFLTGTPDEYVEQVAQYKALPV
      |||||||||||||||||||| |||
MLDAALPPCPTVAATADCEICPAVKRDVDLFLTGTPDEYVEQVAQYNALPVVLENARILKNCVDAKMTEEDK
      10      20  X  30      40      50      60      70
ENALSVLDKIYTSPLC
      80
  
```

3. US-08-300-510-1 (1-27)

JC1136 major allergen chain 1 precursor A - cat

ENTRY JC1136 #type complete
 TITLE major allergen chain 1 precursor A - cat
 ORGANISM #formal_name Felis silvestris catus #common_name domestic cat
 DATE 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change
 31-Dec-1993
 ACCESSIONS JC1136
 REFERENCE JC1126
 #authors Griffith, I.J.; Craig, S.; Pollock, J.; Yu, X.B.;
 Morgenstern, J.P.; Rogers, B.L.
 #journal Gene (1992) 113:263-268
 #title Expression and genomic structure of the genes encoding Fd1,
 the major allergen from the domestic cat.
 #accession JC1136
 ##molecule_type DNA
 ##residues 1-92 ##label GRI
 GENETICS
 #gene Ch1
 #introns 21/1; 83/3
 FEATURE
 1-22 #domain signal sequence #status predicted #label SIG\
 23-92 #product major allergen chain 1 #status predicted #label
 MAT
 SUMMARY #length 92 #molecular-weight 10072 #checksum 4988
 SEQUENCE

Initial Score = 26 Optimized Score = 26 Significance = 16.71
 Residue Identity = 96% Matches = 26 Mismatches = 1
 Gaps = 0 Conservative Substitutions = 0

```

                X      10      20      X
                KRVDLFLTGTPDEYVEQVAQYKALPV
                |||||
MKGACVLVLLWAALLISGGNCEICPAVKRDVDLFLTGTPDEYVEQVAQYNALPVVLENARILKNCVDAKMT
      10      20      30      40      50      X      60      70

EEDKENALSVLDKIYTSPLC
      80      90

```

4. US-08-300-510-1 (1-27)

ISUTTB triose-phosphate isomerase (EC 5.3.1.1) - Trypanos

ENTRY ISUTTB #type complete
 TITLE triose-phosphate isomerase (EC 5.3.1.1) - Trypanosoma brucei
 ALTERNATE_NAMES triosephosphate mutase
 ORGANISM #formal_name Trypanosoma brucei
 DATE 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change
 30-Jun-1993
 ACCESSIONS A25110; A25186
 REFERENCE A25110
 #authors Swinkels, B.W.; Gibson, W.C.; Osinga, K.A.; Kramer, R.;
 Veeneman, G.H.; van Boom, J.H.; Borst, P.
 #journal EMBO J. (1986) 5:1291-1298
 #title Characterization of the gene for the microbody (glycosomal)
 triosephosphate isomerase of Trypanosoma brucei.
 #cross-references MUID:86274631
 #accession A25110
 ##molecule_type DNA
 ##residues 1-250 ##label SWI
 ##cross-references GB:X03921
 COMMENT This enzyme catalyzes the interconversion of glyceraldehyde
 3-phosphate and dihydroxyacetone phosphate.

CLASSIFICATION #superfamily triose-phosphate isomerase
KEYWORDS fatty acid biosynthesis; gluconeogenesis; glycolysis;
homodimer; intramolecular oxidoreductase; isomerase;
pentose phosphate pathway

FEATURE
2-250 #product triose-phosphate isomerase #label MAT\
95,167 #active_site His, Glu #status predicted
SUMMARY #length 250 #molecular-weight 26920 #checksum 4834
SEQUENCE

Initial Score = 10 Optimized Score = 10 Significance = 5.09
Residue Identity = 38% Matches = 10 Mismatches = 16
Gaps = 0 Conservative Substitutions = 0

ACIGETLQERESCR TAVVVL TQIAAI AKKLK KADWAK VVIAYEPV WAI GTGKVATP QQAQEA HALIRSWVSS
130 140 150 160 170 180 190
X 10 20 X
KRDVDLFLTGT PDEYVEQVAQYKALPV
||| ||| ||| |||
KIGADV R GELRIL YGGSVNGKNARTLYQQ RDVNGFLVGGASLKPEFVDI IKATQ
200 210 220 X 230 240 250

5. US-08-300-510-1 (1-27)

PN0644 hypothetical protein 66 - Streptomyces coelicolor

ENTRY PN0644 #type fragment
TITLE hypothetical protein 66 - Streptomyces coelicolor (fragment)
ORGANISM #formal_name Streptomyces coelicolor
DATE 03-May-1994 #sequence_revision 03-May-1994 #text_change
03-May-1994
ACCESSIONS PN0644
REFERENCE JN0831
#authors Wray Jr., L.V.; Fisher, S.H.
#journal Gene (1993) 130:145-150
#title The Streptomyces coelicolor glnR gene encodes a protein
similar to other bacterial response regulators.
#accession PN0644
##molecule_type DNA
##residues 1-66 ##label WRA
##cross-references GB:L03213
GENETICS
#start_codon GTG
SUMMARY #length 66 #checksum 9954
SEQUENCE

Initial Score = 9 Optimized Score = 9 Significance = 4.36
Residue Identity = 33% Matches = 9 Mismatches = 18
Gaps = 0 Conservative Substitutions = 0

X 10 20 X
KRDVDLFLTGT PDEYVEQVAQYKALPV
|| | || | | ||
MAKVTRDDVARLAGTSTAVVSYVINNGPRPVAPATRERVLAAIKELGYRPPDRVAQAMASRRTDLIG
X 10 20 30 40 50 60

6. US-08-300-510-1 (1-27)

LNPG1 pulmonary surfactant protein 9K form - pig

ENTRY LNPG1 #type complete
TITLE pulmonary surfactant protein 9K form - pig
ALTERNATE_NAMES low molecular mass surfactant protein type 1
ORGANISM #formal_name Sus scrofa domestica #common_name domestic pig
DATE 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change

ACCESSIONS S00363
 REFERENCE S00363
 #authors Curstedt, T.; Johansson, J.; Barros-Soederling, J.;
 Robertson, B.; Nilsson, G.; Westberg, M.; Joernvall, H.
 #journal Eur. J. Biochem. (1988) 172:521-525
 #title Low-molecular-mass surfactant protein type 1. The primary
 structure of a hydrophobic 8-kDa polypeptide with eight
 half-cystine residues.
 #cross-references MUID:88166729
 #accession S00363
 ##molecule_type protein
 ##residues 1-79 ##label CUR
 COMMENT Pulmonary surfactant protein is a phospholipid-protein complex,
 which reduces surface tension at the air-liquid interface of the
 alveoli and thus facilitates gaseous exchange.
 CLASSIFICATION #superfamily pulmonary surfactant protein B
 KEYWORDS alveolar proteinosis; gaseous exchange; lipoprotein; lung;
 pulmonary surfactant; respiratory distress syndrome
 SUMMARY #length 79 #molecular-weight 8714 #checksum 5695
 SEQUENCE

Initial Score = 9 Optimized Score = 9 Significance = 4.36
 Residue Identity = 33% Matches = 9 Mismatches = 18
 Gaps = 0 Conservative Substitutions = 0

```

      X      10      20      X
      KRVDLFLTGTPEYVEQVAQYKALPV
      ||      |      ||      ||
FPIPLPFCWLCRTLIKRIQAVVPKGVLLKAVAQVCHVPLPVGGICQCLAERYIVICLNMLLDRTLPLQLVCG
      10      X 20      30      40 X      50      60      70

```

LVLRCSS

7. US-08-300-510-1 (1-27)

S15199 hydrogenase isozyme hypC - Escherichia coli

ENTRY S15199 #type complete
 TITLE hydrogenase isozyme hypC - Escherichia coli
 ORGANISM #formal_name Escherichia coli
 DATE 21-Nov-1993; #sequence_revision 21-Nov-1993; #text_change
 21-Nov-1993
 ACCESSIONS S15199
 REFERENCE S15197
 #authors Lutz, S.; Jacobi, A.; Schlensog, V.; Boehm, R.; Sowers, G.;
 Boeck, A.
 #journal Mol. Microbiol. (1991) 5:123-135
 #title Molecular characterization of an operon (hyp) necessary for
 the activity of the three hydrogenase isoenzymes in
 Escherichia coli.
 #cross-references MUID:91194542
 #accession S15199
 ##status preliminary
 ##residues 1-90 ##label LUT
 ##cross-references EMBL:X54543
 SUMMARY #length 90 #molecular-weight 9732 #checksum 8904
 SEQUENCE

Initial Score = 9 Optimized Score = 12 Significance = 4.36
 Residue Identity = 41% Matches = 14 Mismatches = 13
 Gaps = 7 Conservative Substitutions = 0

```

      X      10      20      X
      KRVDLFLTGTPE----YVEQ---VAQYKALPV

```

MCIGVPGQIRTDGNQAKVDVCGIGRDVDTLVGSCDENGQPRVGQWVLVHVGFMAMSVINEAEARDTLDALQ
 10 20 X 30 40 50 60 70

NMFDVEPDVGALLYGEEK
 80 90

8. US-08-300-510-1 (1-27)
 F53275 kappa 1 b95 allotype=constant region kappa chain -

ENTRY F53275 #type fragment
 TITLE kappa 1 b95 allotype=constant region kappa chain - rabbit
 (fragment)
 ORGANISM #formal_name Oryctolagus cuniculus #common_name domestic
 rabbit
 DATE 02-May-1994 #sequence_revision 18-Nov-1994 #text_change
 18-Nov-1994
 ACCESSIONS F53275
 REFERENCE A53275
 #authors Ayadi, H.; Marche, P.N.; Cazenave, P.A.
 #journal Immunogenetics (1991) 34:201-207
 #title Evolution of the rabbit immunoglobulin kappa chain genes.
 #cross-references MUID:91372868
 #accession F53275
 ##status preliminary
 ##molecule_type DNA
 ##residues 1-104 ##label AYA
 ##cross-references NCBIN:56069; NCBIP:56170
 ##note sequence inconsistent with nucleotide translation
 ##note sequence extracted from NCBI backbone
 SUMMARY #length 104 #checksum 7726
 SEQUENCE

Initial Score = 9 Optimized Score = 11 Significance = 4.36
 Residue Identity = 34% Matches = 12 Mismatches = 15
 Gaps = 8 Conservative Substitutions = 0

X 10
 KRVDLFLTGT-
 | | |
 DPVAPTVLIFPPSPAELATGTATIVCVANKYFPDVTVTWKVDGTTQTTGIENSRTFPQNSDDCTYNLSSTLTL
 10 20 30 40 50 60 70
 20 X
 -----PDEYVEQVAGYKALPV
 ||| ||| ||
 KSDEYNHDEYICQVAGGSGSPVVQSF SRNNC
 80 90 X 100

9. US-08-300-510-1 (1-27)
 K5RBV Ig kappa chain C region (B5 variant) - rabbit

ENTRY K5RBV #type complete
 TITLE Ig kappa chain C region (B5 variant) - rabbit
 ORGANISM #formal_name Oryctolagus cuniculus #common_name domestic
 rabbit
 DATE 13-Aug-1986 #sequence_revision 13-Aug-1986 #text_change
 04-Nov-1994
 ACCESSIONS A02124
 REFERENCE A02124
 #authors Bernstein, K.E.; Skurla Jr., R.M.; Mage, R.G.
 #journal Nucleic Acids Res. (1983) 11:7205-7214
 #title The sequences of rabbit kappa light chains of b4 and b5
 allotypes differ more in their constant regions than in

their 3 untranslated regions.

#cross-references MUID:84041515

#contents Clone pkb5-F2

#accession A02124

##molecule_type mRNA

##residues 1-104 ##label BER

##note the cDNA from which this sequence was derived contains a terminator codon within the V-region coding region; the origin of this codon and of the differences between this and other sequenced b5 C regions are unclear; the cDNA clone was made using mRNA from trypanosome-infected b5-homozygous rabbits

CLASSIFICATION #superfamily immunoglobulin C region; immunoglobulin homology

KEYWORDS immunoglobulin

FEATURE

1-104 #domain C region #label CRG\

19-87 #domain immunoglobulin homology #label IMM

SUMMARY #length 104 #molecular-weight 11079 #checksum 6706

SEQUENCE

Initial Score = 9 Optimized Score = 11 Significance = 4.36
Residue Identity = 34% Matches = 12 Mismatches = 15
Gaps = 8 Conservative Substitutions = 0

```

                                     X      10
                                     KRVDLFLTGT-
                                     |  |  |
ATLAPTVLIFPPSPAELATGTATIVCVANKYFPDGTVTWQVDGKPLTTGIETSKTPQNSDDCTYNLSSTLTL
      10      20      30      40      50      60      70

      20      X
-----PDEYVEQVAQYKALPV
      |||  |||  ||
KSDEYNSHDEYTCQVAQSGSPVVSFSRKNC
      80      90      X  100
```

10. US-08-300-510-1 (1-27)

A20968 Ig kappa-1b5 chain C region - rabbit (fragment)

ENTRY A20968 #type fragment

TITLE Ig kappa-1b5 chain C region - rabbit (fragment)

ORGANISM #formal_name *Oryctolagus cuniculus* #common_name domestic rabbit

DATE 10-Aug-1990 #sequence_revision 10-Aug-1990 #text_change 23-Mar-1993

ACCESSIONS A20968

REFERENCE A20968

#authors Emorine, L.; Sognr, J.A.; Trinh, D.; Kindt, T.J.; Max, E.E.

#journal Proc. Natl. Acad. Sci. U.S.A. (1984) 81:1789-1793

#title A genomic gene encoding the b5 rabbit immunoglobulin kappa constant region: implications for latent allotype phenomenon.

#cross-references MUID:84170387

#accession A20968

##status preliminary

##molecule_type DNA

##residues 1-105 ##label EMO

SUMMARY #length 105 #checksum 237

SEQUENCE

Initial Score = 9 Optimized Score = 11 Significance = 4.36
Residue Identity = 34% Matches = 12 Mismatches = 15
Gaps = 8 Conservative Substitutions = 0

X 10

VATLAPTVLIFPPSPAELATGTATIVCVANKYFPDGTVTWQVDGKPLTTGIETSKTPQNSDDCTYNLSSTLT
 10 20 30 40 50 60 X 70

-----PDEYVEQVAQYKALPV
 LKSDEYNHSHDEYTCQVAQSGSPVVQSF SRKNC
 80 90 X 100

11. US-08-300-510-1 (1-27)
 S43188 orotidine-5'-phosphate decarboxylase (EC 4.1.1.23)

ENTRY S43188 #type complete
 TITLE orotidine-5'-phosphate decarboxylase (EC 4.1.1.23) -
 Pseudomonas aeruginosa
 ORGANISM #formal_name Pseudomonas aeruginosa
 DATE 20-May-1994; #sequence_revision 20-May-1994; #text_change
 20-May-1994
 ACCESSIONS S43188
 REFERENCE S43188
 #authors Strych, U.; Wohlfarth, S.; Winkler, U.K.
 #submission submitted to the EMBL Data Library, April 1992
 #accession S43188
 ##status preliminary
 ##residues 1-232 ##label STR
 ##cross-references EMBL:X65613
 SUMMARY #length 232 #molecular-weight 24368 #checksum 1180
 SEQUENCE

Initial Score = 9 Optimized Score = 9 Significance = 4.36
 Residue Identity = 33% Matches = 9 Mismatches = 18
 Gaps = 0 Conservative Substitutions = 0

ADQLDPKLCRVKVGKELFTSCAAGIVETLRGKGFEVFLDLKFHDIPNTTAMAVKAAAEMGVMMVNVHCSGGL
 30 40 50 60 70 80 90

X 10 20 X
 KRQVDLFLTGTDPDEYVEQVAQYKALPV
 RMMAACRETLEAFSGARPLLIGVTVLTSMEREDLAGIGLDIEPQEQVLRRLAALAQKAGMDGLVCSAQEAPAL
 100 110 120 X 130 140 150 160

KAAHPGLQLVTPGIRPAGSAQDDQRRILTPRQALDAGSDYLVIGRPISQAADPAKALAAIIVAEELG
 170 180 190 200 210 220 230

12. US-08-300-510-1 (1-27)
 JS0618 glutathione transferase (EC 2.5.1.18) Yrs precursor

ENTRY JS0618 #type complete
 TITLE glutathione transferase (EC 2.5.1.18) Yrs precursor - rat
 ORGANISM #formal_name Rattus norvegicus #common_name Norway rat
 DATE 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change
 18-Jun-1993
 ACCESSIONS JS0618; PS0266
 REFERENCE JS0618
 #authors Ogura, K.; Nishiyama, T.; Okada, T.; Kajita, J.; Narihata,
 H.; Watabe, T.; Hiratsuka, A.; Watabe, T.
 #journal Biochem. Biophys. Res. Commun. (1991) 181:1294-1300
 #title Molecular cloning and amino acid sequencing of rat liver
 class theta glutathione S-transferase Yrs-Yrs inactivating
 reactive sulfate esters of carcinogenic arylmethanols.
 #cross-references MUID:92109741


```

#accession 030610
##molecule_type mRNA
##residues 1-244 ##label OGU
#accession PS0266
##molecule_type protein
##residues 2-34;38-51;93-114;140-159;195-209;215-228 ##label OGU1
COMMENT Glutathione transferase Yrs-Yrs is composed of two identical
chains.
COMMENT Glutathione transferase Yrs-Yrs belongs to class theta.
KEYWORDS transferase
FEATURE
2-244 #product glutathione transferase Yrs chain #status
experimental #label GLU
SUMMARY #length 244 #molecular-weight 27439 #checksum 3952
SEQUENCE

```

```

Initial Score = 9 Optimized Score = 10 Significance = 4.36
Residue Identity = 37% Matches = 10 Mismatches = 17
Gaps = 0 Conservative Substitutions = 0

```

```

          X      10      20      X
          KRVDLFLTGTPEYVEQVAQYKALPV
          | || | | || | | |
MGLELYLDLLSQPSRAVYIFAKKNGIPFQLRTVDLLKGQHLSEQFSQVNCCLKKVPVLKDGSFVLTESTAILI
      10      20      30      40      50      X 60      70
YLSSKYQVADHWYPADLQARAQVHEYLGNHADNIRGTFGVLLWTKVLGPLIGVQVPPEKVERNRNSMVLALQ
      80      90     100     110     120     130     140
RLEDKFLRDRAF
      150

```

13. US-08-300-510-1 (1-27)
A38233 triose-phosphate isomerase (EC 5.3.1.1) - fluke (S

```

ENTRY A38233 #type complete
TITLE triose-phosphate isomerase (EC 5.3.1.1) - fluke (Schistosoma
mansoni)
ALTERNATE_NAMES triosephosphate mutase
ORGANISM #formal_name Schistosoma mansoni
DATE 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change
31-Dec-1993
ACCESSIONS A38233
REFERENCE A38233
#authors Shoemaker, C.; Gross, A.; Gebremichael, A.; Harn, D.
#journal Proc. Natl. Acad. Sci. U.S.A. (1992) 89:1842-1846
#title cDNA cloning and functional expression of the Schistosoma
mansoni protective antigen triose-phosphate isomerase.
#cross-references MUID:92179278
#accession A38233
##molecule_type mRNA
##residues 1-253 ##label SHO
##cross-references NCBIP:87225
##note sequence extracted from NCBI backbone
CLASSIFICATION #superfamily triose-phosphate isomerase
KEYWORDS fatty acid biosynthesis; gluconeogenesis; glycolysis;
homodimer; intramolecular oxidoreductase; isomerase;
pentose phosphate pathway
SUMMARY #length 253 #molecular-weight 28122 #checksum 3727
SEQUENCE

```

```

Initial Score = 9 Optimized Score = 9 Significance = 4.36
Residue Identity = 33% Matches = 9 Mismatches = 18
Gaps = 0 Conservative Substitutions = 0

```

CIGETLSERESNRTEEVGVRLRAIANKIKSADDEWRRVVVAYEPVWAIGTGLVAIFQGMQEVANFLAAWPA
 130 140 150 160 170 180 190

X 10 20 X
 KRVDVDFLTGTPDEYVEQVAQYKALPV
 ||| ||| | ||
 NAPNGVDEKIRIIYGGSVTAANCKELAQGHVDGFLVGGASLKPEFTEICKARQR
 200 210 220 230 240 250 X

14. US-08-300-510-1 (1-27)

S04405 hydroxyneurosporene synthase - Rhodobacter capsula

ENTRY S04405 #type complete
 TITLE hydroxyneurosporene synthase - Rhodobacter capsulatus
 ORGANISM #formal_name Rhodobacter capsulatus
 DATE 28-Feb-1990 #sequence_revision 28-Feb-1990 #text_change
 18-Jun-1993
 ACCESSIONS S04405
 REFERENCE S04401
 #authors Armstrong, G.A.; Alberti, M.; Leach, F.; Hearst, J.E.
 #journal Mol. Gen. Genet. (1989) 216:254-268
 #title Nucleotide sequence, organization, and nature of the protein
 products of the carotenoid biosynthesis gene cluster of
 Rhodobacter capsulatus.
 #cross-references MUID:89313663
 #accession S04405
 ##molecule_type DNA
 ##residues 1-281 ##label ARM
 ##cross-references EMBL:X52291
 GENETICS
 #gene crtC
 KEYWORDS carotenoid biosynthesis
 SUMMARY #length 281 #molecular-weight 31856 #checksum 8228
 SEQUENCE

Initial Score	=	9	Optimized Score	=	10	Significance	=	4.36
Residue Identity	=	35%	Matches	=	11	Mismatches	=	16
Gaps	=	4	Conservative Substitutions	=			=	0

X 10
 KRD----VDLFLTGTGTPDEYVEQ
 || ||| |||
 MIAFIGSVFSPWYRWSGRREPQNHCCINMVTGTGDRFTMTDRGRSALRQSRDSFGVGP SKLTWTGKELVID
 10 20 30 40 50 60 70
 20 X
 VAQYKALPV
 | |||
 VDEWGALPKLGKLGKRVVLTTPRAVTGVEVRLTPDAGHTWRPFAPADVEVDLAPGHKWTGHGYFDANFGTRA
 80 90 100 110 120 130 140
 LEEDFSFWTGRFPLKDRTVCFYDATRLDRTKVALAV
 150 160 170 180

15. US-08-300-510-1 (1-27)

S21394 transposase - Mycobacterium tuberculosis

ENTRY S21394 #type complete
 TITLE transposase - Mycobacterium tuberculosis
 ORGANISM #formal_name Mycobacterium tuberculosis
 DATE 22-Nov-1993; #sequence_revision 22-Nov-1993; #text_change
 22-Nov-1993
 ACCESSIONS S21394
 REFERENCE S21394

#authors Mariani, F.; Piccollella, E.; Colizzi, V.; Nappucci, R.; Gross,
R.
#submission submitted to the EMBL Data Library, April 1992
#accession S21394
##status preliminary
##residues 1-308 ##label MAR
##cross-references EMBL:X65618
SUMMARY #length 308 #molecular-weight 34272 #checksum 8922
SEQUENCE

Initial Score = 9 Optimized Score = 10 Significance = 4.36
Residue Identity = 30% Matches = 12 Mismatches = 15
Gaps = 12 Conservative Substitutions = 0

```

                                     X
                                     KRVDLFL
                                     ||  |
MTRVGVISDEFWAVVEPLMPSSHEGKPGRRFSDHRLILEGIAWRFRGTGSPWRDLPAEFGPWQT VWKRRHHRWSL
      10      20      30      40      50      60      X      70

      10      20      X
      TGTPDEYVEQVA-----GYKALPV
      || ||  ||      | | |
      DGTCDDEVFAHVAAVFGVDAEVAEDIEKLLSVDSTNVRAHQHSAGAARTRSPQALSDYKKSADPPDDHAIGR
      80      90      100 X      110      120      130      140

      SRGGLTTKIHALTDQREAPVRIRLTAGQAGDNPQLPLLDYRHASTEYALGSTDFRL
      150      160      170      180      190      200

```

> 0 <
0| |0 IntelliGenetics
> 0 <

FastDB - Fast Pairwise Comparison of Sequences
Release 5.4

Results file 1-spt.res made by on Fri 24 Mar 95 7:53:44-PST.

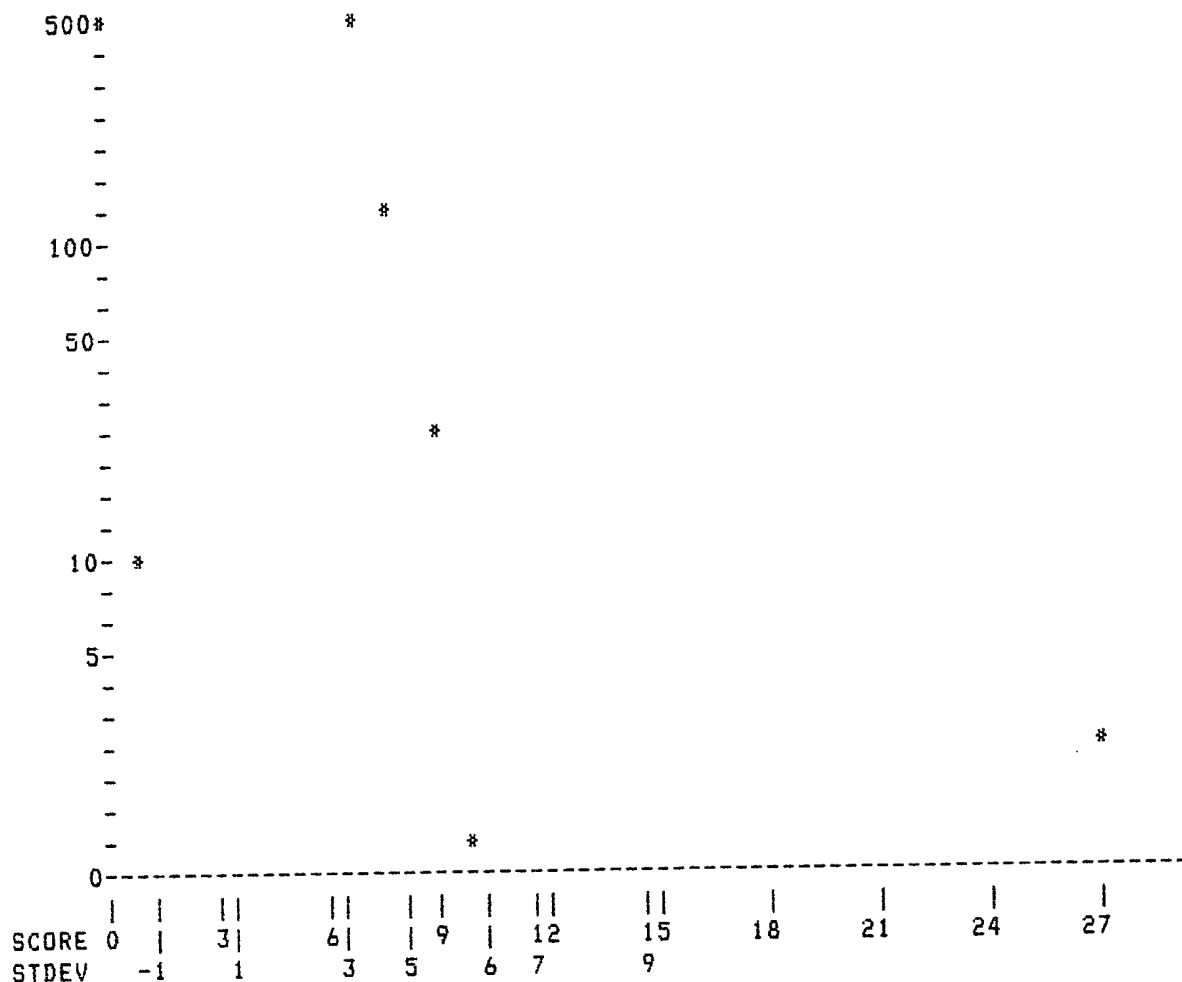
Query sequence being compared: US-08-300-510-1 (1-27)
Number of sequences searched: 40292
Number of scores above cutoff: 3849

Results of the initial comparison of US-08-300-510-1 (1-27) with:
Data bank : Swiss-Prot 30, all entries

```

100000-
-
N -
U50000-
M -
B -
E -
R -
-
D - *
F10000- * *
-
S - *
E 5000-
Q -
U -
E - *
N -
C -
E -
S 1000-

```



PARAMETERS

Similarity matrix	Unitary	K-tuple	2
Mismatch penalty	1	Joining penalty	20
Gap penalty	1.00	Window size	27
Gap size penalty	0.05		
Cutoff score	0		
Randomization group	0		
Initial scores to save	40	Alignments to save	15
Optimized scores to save	0	Display context	100

SEARCH STATISTICS

Scores:	Mean	Median	Standard Deviation
	3	5	1.31

Times:	CPU	Total Elapsed
	00:00:38.04	00:00:40.00

Number of residues:	14147368
Number of sequences searched:	40292
Number of scores above cutoff:	3849

Cut-off raised to 3.
 Cut-off raised to 4.
 Cut-off raised to 5.
 Cut-off raised to 6.
 Cut-off raised to 7.

The scores below are sorted by initial score.

Significance is calculated based on initial score.

2 100% similar sequences to the query sequence were found:

Sequence Name	Description	Length	Init. Score	Opt. Score	Sig.	Frame
1. FELA_FELCA	MAJOR ALLERGEN I POLYPEPTIDE	92	27	27	18.34	0
2. FELB_FELCA	MAJOR ALLERGEN I POLYPEPTIDE	88	27	27	18.34	0

The list of other best scores is:

Sequence Name	Description	Length	Init. Score	Opt. Score	Sig.	Frame
3. TPIS_TRYBB	**** 5 standard deviations above mean **** TRIOSEPHOSPHATE ISOMERASE, GL	250	10	10	5.35	0
4. PSPB_PIG	**** 4 standard deviations above mean **** PULMONARY SURFACTANT-ASSOCIAT	79	9	9	4.58	0
5. HYPC_ECOLI	HYDROGENASE ISOENZYMES FORMAT	90	9	12	4.58	0
6. KAC6_RABIT	IG KAPPA CHAIN B5 VARIANT C R	104	9	11	4.58	0
7. FIMA_BORPE	FIMA PROTEIN.	145	9	11	4.58	0
8. GTTR_RAT	GLUTATHIONE S-TRANSFERASE YRS	243	9	10	4.58	0
9. CRTC_RHOCA	HYDROXYNEUROSPORENE DEHYDROGE	281	9	10	4.58	0
10. FVT1_HUMAN	FOLLICULAR VARIANT TRANSLOCAT	332	9	9	4.58	0
11. E13B_HORVU	GLUCAN ENDO-1,3-BETA-GLUCOSID	334	9	9	4.58	0
12. ALGP_PSEAE	TRANSCRIPTIONAL REGULATORY PR	340	9	9	4.58	0
13. CHLI_EUGGR	PROBABLE MAGNESIUM-CHELATASE	348	9	9	4.58	0
14. PRTZ_HORVU	PROTEIN Z (Z4) (MAJOR ENDOSPE	399	9	9	4.58	0
15. PRE2_STAAU	PLASMID RECOMBINATION ENZYME	420	9	9	4.58	0
16. CHLI_ARATH	PROBABLE MAGNESIUM-CHELATASE	424	9	9	4.58	0
17. ARDA_MYCTU	3-PHOSPHOSHIKIMATE 1-CARBOXYV	450	9	9	4.58	0
18. PCD_ARTOX	PHENMEDIPHAM HYDROLASE (3.1.1	493	9	11	4.58	0
19. AMPL_SOLTU	CYTOSOL AMINOPEPTIDASE (EC 3.	554	9	10	4.58	0
20. GAG_SFV1	GAG POLYPROTEIN (CORE POLYPRO	647	9	10	4.58	0
21. PUPB_PSEPU	FERRIC-PSEUDOBACTIN RECEPTOR	809	9	9	4.58	0
22. PTF1_RHOCA	MULTIPHOSPHORYL TRANSFER PROT	827	9	10	4.58	0
23. VIRA_AGRT9	WIDE HOST RANGE (WHR) VIRA PR	829	9	9	4.58	0
24. VIRA_AGRT6	WIDE HOST RANGE (WHR) VIRA PR	829	9	9	4.58	0
25. IC18_HCMVA	PROBABLE PROCESSING AND TRANS	850	9	10	4.58	0
26. RPB2_SCHPD	DNA-DIRECTED RNA POLYMERASE I	1210	9	10	4.58	0
27. MYSC_CAEEL	MYOSIN HEAVY CHAIN C (MHC C).	1947	9	9	4.58	0
28. RL27_PEA	**** 3 standard deviations above mean **** 60S RIBOSOMAL PROTEIN L27.	135	8	9	3.82	0
29. RL25_YEAST	60S RIBOSOMAL PROTEIN L25 (RP	136	8	9	3.82	0
30. HUPG_RHILV	HUPG PROTEIN.	149	8	8	3.82	0
31. PCCB_HUMAN	PROPIONYL-COA CARBOXYLASE BET	155	8	9	3.82	0
32. MEMG_METTR	METHANE MONOOXYGENASE COMPONE	169	8	8	3.82	0
33. YGT2_CHLPS	HYPOTHETICAL 21.0 KD PROTEIN	182	8	9	3.82	0
34. YC08_YEAST	HYPOTHETICAL 21.1 KD PROTEIN	192	8	8	3.82	0
35. YEGA_ECOLI	HYPOTHETICAL IN DCD 3'REGION	200	8	8	3.82	0
36. COX3_THEP3	CYTOCHROME C OXIDASE POLYPEPT	207	8	8	3.82	0
37. SSPA_ECOLI	STRINGENT STARVATION PROTEIN	212	8	9	3.82	0
38. SI21_RAT	SERINE PROTEASE INHIBITOR 2.1	214	8	11	3.82	0
39. VHEL_LVX	PROBABLE HELICASE (ORF 2).	216	8	9	3.82	0
40. YIAQ_ECOLI	HYPOTHETICAL 23.4 KD PROTEIN	220	8	9	3.82	0

1. US-08-300-510-1 (1-27)

FELA_FELCA MAJOR ALLERGEN I POLYPEPTIDE CHAIN 1 MAJOR FORM PR

ID FELA_FELCA STANDARD; PRT: 92 AA.
AC P30438;
DT 01-APR-1993 (REL. 25, CREATED)
DT 01-APR-1993 (REL. 25, LAST SEQUENCE UPDATE)

DT 01-JUN-1994 (REL. 27, LAST ANNOTATION UPDATE)
 DE MAJOR ALLERGEN I POLYPEPTIDE CHAIN 1 MAJOR FORM PRECURSOR (FEL D 1)
 DE (CAT-1) (AG 4).
 GN CH1.
 OS FELIS CATUS (CAT).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
 OC EUTHERIA; CARNIVORA.
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 23-92.
 RC TISSUE=SALIVARY GLAND;
 RM 92052157
 RA MORGENSTERN J.P., GRIFFITH I.J., BRAUER A.W., ROGERS B.L.,
 RA BOND J.F., CHAPMAN M.D., KUO M.-C.;
 RL PROC. NATL. ACAD. SCI. U.S.A. 88:9690-9694(1991).
 RN [2]
 RP SEQUENCE FROM N.A.
 RM 92241678
 RA GRIFFITH I.J., CRAIG S., POLLOCK J., YU X.-B., MORGENSTERN J.P.,
 RA ROGERS B.L.,
 RL GENE 113:263-268(1992).
 RN [3]
 RP SEQUENCE OF 23-62, AND CHARACTERIZATION.
 RM 91287714
 RA DUFFORT D.A., CARREIRA J., NITTI G., POLO F., LOMBARDERO M.;
 RL MOL. IMMUNOL. 28:301-309(1991).
 RN [4]
 RP CHARACTERIZATION.
 RA LEITERMANN K., OHMAN J.L. JR.;
 RL J. ALLERGY CLIN. IMMUNOL. 74:147-153(1991).
 CC -!- DISEASE: MAJOR ALLERGEN PRODUCED BY THE DOMESTIC CAT.
 CC -!- SUBUNIT: HETEROTETRAMER COMPOSED OF TWO NON-COVALENTLY LINKED
 CC DISULFIDE-LINKED HETERODIMER OF CHAINS 1 AND 2.
 CC -!- TISSUE SPECIFICITY: SALIVA, AND SEBACEOUS GLANDS.
 CC -!- ALTERNATIVE PRODUCTS: USAGE OF TWO DIFFERENT INITIATOR MET ARE
 CC RESPONSIBLE FOR THE PRODUCTION OF TWO FORMS OF THE SIGNAL SEQUENCE
 CC OF THIS ALLERGEN SUBUNIT.
 CC -!- SIMILARITY: TO UTEROGLOBIN.
 DR EMBL; M74952; FDFELD1.
 DR PIR; JC1136; JC1136.
 DR PROSITE; PS00403; UTEROGLOBIN_1.
 DR PROSITE; PS00404; UTEROGLOBIN_2.
 KW ALLERGEN; SIGNAL; ALTERNATIVE SPLICING.
 FT SIGNAL 1 22
 FT CHAIN 23 92 MAJOR ALLERGEN I POLYPEPTIDE CHAIN 1.
 FT DISULFID 25 25 INTERCHAIN (POTENTIAL).
 FT DISULFID 92 92 INTERCHAIN (POTENTIAL).
 FT VARIANT 51 51 K -> N.
 FT CONFLICT 5 5 R -> C (IN REF. 2).
 FT CONFLICT 18 18 W -> S (IN REF. 2).
 FT CONFLICT 82 82 L -> V (IN REF. 2).
 SQ SEQUENCE 92 AA; 10252 MW; 43206 CN;

Initial Score = 27 Optimized Score = 27 Significance = 18.34
 Residue Identity = 100% Matches = 27 Mismatches = 0
 Gaps = 0 Conservative Substitutions = 0

X 10 20 X
 KRVDLFLTGTPDEYVEQVAGYKALPV
 ||||||||||||||||||||
 MKGARVLVLLWAALLLIWGNCEICPAVKRVDLFLTGTPDEYVEQVAGYKALPVVLENARILKNCVDAMT
 10 20 30 40 50 X 60 70
 EEDKENALSLLDKIYTSPLC
 80 90

FELB_FELCA MAJOR ALLERGEN I POLYPEPTIDE CHAIN 1 MINOR FORM PR

ID FELB_FELCA STANDARD; PRT; 88 AA.
AC P30439;
DT 01-APR-1993 (REL. 25, CREATED)
DT 01-APR-1993 (REL. 25, LAST SEQUENCE UPDATE)
DT 01-JUN-1994 (REL. 29, LAST ANNOTATION UPDATE)
DE MAJOR ALLERGEN I POLYPEPTIDE CHAIN 1 MINOR FORM PRECURSOR (FEL D 1)
DE (CAT-1) (AG 4).
GN CH1.
OS FELIS CATUS (CAT).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; CARNIVORA.
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 19-88.
RM 92052157
RA MORGENSTERN J.P., GRIFFITH I.J., BRAUER A.W., ROGERS B.L.,
RA BOND J.F., CHAPMAN M.D., KUO M.-C.;
RL PROC. NATL. ACAD. SCI. U.S.A. 88:9690-9694(1991).
RN [2]
RP SEQUENCE FROM N.A.
RM 92241678
RA GRIFFITH I.J., CRAIG S., POLLOCK J., YU X.-B., MORGENSTERN J.P.,
RA ROGERS B.L.,
RL GENE 113:263-268(1992).
RN [3]
RP SEQUENCE OF 19-58, AND CHARACTERIZATION.
RM 91287714
RA DUFFORT D.A., CARREIRA J., NITTI G., POLO F., LOMBARDERO M.;
RL MDL. IMMUNOL. 28:301-309(1991).
RN [4]
RP CHARACTERIZATION.
RA LEITERMANN K., OHMAN J.L. JR.;
RL J. ALLERGY CLIN. IMMUNOL. 74:147-153(1991).
CC -!- DISEASE: MAJOR ALLERGEN PRODUCED BY THE DOMESTIC CAT.
CC -!- SUBUNIT: HETEROTETRAMER COMPOSED OF TWO NON-COVALENTLY LINKED
CC DISULFIDE-LINKED HETERODIMER OF CHAINS 1 AND 2.
CC -!- TISSUE SPECIFICITY: SALIVA, AND SEBACEOUS GLANDS.
CC -!- ALTERNATIVE PRODUCTS: USAGE OF TWO DIFFERENT INITIATOR MET ARE
CC RESPONSIBLE FOR THE PRODUCTION OF TWO FORMS OF THE SIGNAL SEQUENCE
CC OF THIS ALLERGEN SUBUNIT.
CC -!- SIMILARITY: TO UTEROGLOBIN.
DR EMBL; M74953; FDFELDIB.
DR PIR; JC1126; JC1126.
DR PROSITE; PS00403; UTEROGLOBIN_1.
DR PROSITE; PS00404; UTEROGLOBIN_2.
KW ALLERGEN; SIGNAL; ALTERNATIVE SPLICING.
FT SIGNAL 1 18
FT CHAIN 19 88 MAJOR ALLERGEN I POLYPEPTIDE CHAIN 1.
FT DISULFID 21 21 INTERCHAIN (POTENTIAL).
FT DISULFID 88 88 INTERCHAIN (POTENTIAL).
FT VARIANT 47 47 K -> N.
FT CONFLICT 78 78 L -> V (IN REF. 2).
SQ SEQUENCE 88 AA; 9614 MW; 39445 CN;

Initial Score = 27 Optimized Score = 27 Significance = 18.34
Residue Identity = 100% Matches = 27 Mismatches = 0
Gaps = 0 Conservative Substitutions = 0

X 10 20 X
KRDVDLFLTGTPEYVEQVAQYKALPV
|||||
MLDAALPPCPTVAATADCEICPAVKRDVDLFLTGTPEYVEQVAQYKALPVVLENARILKNCVDAKMTEEDK
10 20 X 30 40 50 60 70

3. US-08-300-510-1 (1-27)

TPIS_TRYBB TRIOSEPHOSPHATE ISOMERASE, GLYCOSOMAL (EC 5.3.1.1)

ID TPIS_TRYBB STANDARD; PRT: 250 AA.
AC P04789;
DT 13-AUG-1987 (REL. 05, CREATED)
DT 13-AUG-1987 (REL. 05, LAST SEQUENCE UPDATE)
DT 01-JUN-1994 (REL. 29, LAST ANNOTATION UPDATE)
DE TRIOSEPHOSPHATE ISOMERASE, GLYCOSOMAL (EC 5.3.1.1) (TIM).
OS TRYPAOSOMA BRUCEI BRUCEI.
OC EUKARYOTA; PROTOZOA; SARCOMASTIGOPHORA; MASTIGOPHORA; KINETOPLASTIDA;
OC TRYPAOSOMATIDAE.
RN [1]
RP SEQUENCE FROM N.A.
RM 86274631
RA SWINKELS B.W., GIBSON W.C., OSINGA K.A., KRAMER R., VEENEMAN G.H.,
RA VAN BOOM J.H., BORST P.;
RL EMBO J. 5:1291-1298(1986).
RN [2]
RP SEQUENCE.
RM 86187863
RA BORST P.;
RL BIOCHIM. BIOPHYS. ACTA 866:179-203(1986).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS).
RM 88118904
RA WIERENGA R.K., KALK K.H., HOL W.J.G.;
RL J. MOL. BIOL. 198:109-121(1987).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (1.83 ANGSTROMS).
RM 91350193
RA WIERENGA R.K., NOBLE M.E.M., VRIEND G., NAUCHE S., HOL W.J.G.;
RL J. MOL. BIOL. 220:995-1015(1991).
RN [5]
RP X-RAY CRYSTALLOGRAPHY.
RM 92235847
RA WIERENGA R.K., NOBLE M.E.M., DAVENPORT R.C.;
RL J. MOL. BIOL. 224:1115-1126(1992).
CC -!- CATALYTIC ACTIVITY: D-GLYCERALDEHYDE 3-PHOSPHATE = DIHYDROXY-
CC ACETONE PHOSPHATE.
CC -!- SUBUNIT: HOMODIMER.
CC -!- PATHWAY: PLAYS AN IMPORTANT ROLE IN SEVERAL METABOLIC PATHWAYS.
CC -!- SUBCELLULAR LOCATION: GLYCOSOMAL.
CC -!- THE ENZYME CONTAINS A HIGH PROPORTION OF POSITIVELY-CHARGED
CC RESIDUES IN BETA-BARRELS V & VII (COMPARED TO THE HOMOLOGOUS
CC REGIONS IN OTHER TRIOSE ISOMERASE SEQUENCES). SINCE 2 CLUSTERS
CC OF + CHARGES LOCATED AT PRECISE DISTANCES ON THE MOLECULAR SURFACE
CC ARE COMMON TO 4 GLYCOSOMAL ENZYMES, [1] SPECULATES THAT THIS MIGHT
CC REPRESENT A SIGNAL FOR ENTRY INTO GLYCOSOMES.
DR EMBL: X03921; TBTIM.
DR PIR: A25110; ISUTTB.
DR PDB: 3TIM; 15-OCT-91.
DR PDB: 4TIM; 15-OCT-92.
DR PDB: 5TIM; 15-OCT-92.
DR PDB: 6TIM; 31-JAN-94.
DR PDB: 1TRD; 31-OCT-93.
DR PDB: 1TSI; 31-JAN-94.
DR PROSITE: PS00171; TIM.
KW ISOMERASE; GLYCOLYSIS; GLUCONEOGENESIS; FATTY ACID BIOSYNTHESIS;
KW PENTOSE SHUNT; GLYCOSOME; 3D-STRUCTURE.
FT ACT_SITE 95 95
FT ACT_SITE 167 167

FT	STRAND	7	11
FT	STRAND	14	14
FT	HELIX	18	30
FT	STRAND	38	43
FT	TURN	46	47
FT	HELIX	48	54
FT	TURN	58	59
FT	STRAND	60	64
FT	STRAND	68	68
FT	STRAND	72	72
FT	TURN	75	76
FT	STRAND	79	79
FT	HELIX	80	85
FT	TURN	86	87
FT	STRAND	90	93
FT	HELIX	96	101
FT	HELIX	106	118
FT	TURN	119	120
FT	STRAND	122	127
FT	HELIX	131	135
FT	TURN	136	137
FT	HELIX	139	151
FT	TURN	152	153
FT	HELIX	156	161
FT	STRAND	162	166
FT	HELIX	169	171
FT	HELIX	180	197
FT	TURN	198	198
FT	HELIX	200	205
FT	STRAND	207	210
FT	HELIX	216	223
FT	TURN	224	224
FT	TURN	226	227
FT	STRAND	230	233
FT	HELIX	235	238
FT	TURN	240	241
FT	HELIX	242	247
FT	TURN	248	249
SO	SEQUENCE	250 AA;	26920 MW; 313102 CN;

Initial Score = 10 Optimized Score = 10 Significance = 5.35
 Residue Identity = 38% Matches = 10 Mismatches = 16
 Gaps = 0 Conservative Substitutions = 0

ACIGETLQERESGRTAVVVL	TQIAAIAKKLKKADWAKV	VIAYEPVWAIGTGK	VATPQQAQEAHALIRSW	VSS
130	140	150	160	170 180 190
		X	10	20 X
		KRDVDLFLTGTPDEYVEQ	VAQYKALPV	
KIGADVRGELRILYGGSV	NGKNARTLYQQRDVNG	FLVGGASLKPEFVDII	KATQ	
200	210	220 X	230 240 250	

4. US-08-300-510-1 (1-27)

PSPB_PIG PULMONARY SURFACTANT-ASSOCIATED PROTEIN B (SP-B) (

ID PSPB_PIG STANDARD; PRT; 79 AA.
 AC P15782;
 DT 01-APR-1990 (REL. 14, CREATED)
 DT 01-APR-1990 (REL. 14, LAST SEQUENCE UPDATE)
 DT 01-JUN-1994 (REL. 29, LAST ANNOTATION UPDATE)
 DE PULMONARY SURFACTANT-ASSOCIATED PROTEIN B (SP-B) (8 KD PROTEIN)
 DE (PULMONARY SURFACTANT-ASSOCIATED PROTEOLIPID SPL(PHE)).
 OS SUS SCROFA (PIG).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
 CC

CC ESTHERIA; HYDROLYTIC.
 RN [1]
 RP SEQUENCE.
 RM 88166729
 RA CURSTEDT T., JOHANSSON J., BARROS-SOEDERLING J., ROBERTSON B.,
 RA NILSSON G., WESTBERG M., JOERNVALL H.;
 RL EUR. J. BIOCHEM. 172:521-525(1988).
 RN [2]
 RP DISULFIDE BONDS.
 RM 91299745
 RA JOHANSSON J., CURSTEDT T., JOERNVALL H.;
 RL BIOCHEMISTRY 30:6917-6921(1991).
 CC -!- FUNCTION: PULMONARY SURFACTANT ASSOCIATED PROTEINS PROMOTE
 CC ALVEOLAR STABILITY BY LOWERING THE SURFACE TENSION AT THE AIR-
 CC LIQUID INTERFACE IN THE PERIPHERAL AIR SPACES. SP-B INCREASES
 CC THE COLLAPSE PRESSURE OF PALMITIC ACID TO NEARLY 70 MILLINEWTONS
 CC PER METER.
 CC -!- PULMONARY SURFACTANT CONSISTS OF 90% LIPID AND 10% PROTEIN. THERE
 CC ARE 4 SURFACTANT ASSOCIATED PROTEIN: 2 COLLAGENOUS, CARBOHYDRATE-
 CC BINDING GLYCOPROTEINS (SP-A AND SP-D) AND 2 SMALL HYDROPHOBIC
 CC PROTEINS (SP-B AND SP-C).
 CC -!- SUBCELLULAR LOCATION: EXTRACELLULAR.
 CC -!- SUBUNIT: HOMODIMER, DISULFIDE-LINKED.
 DR PIR: S00363; LNPG1.
 KW SURFACE FILM; LUNG; GASEOUS EXCHANGE.
 FT DISULFID 8 77
 FT DISULFID 11 71
 FT DISULFID 35 46
 FT DISULFID 48 48 INTERCHAIN.
 FT VARIANT 57 57 C -> L.
 SQ SEQUENCE 79 AA; 8714 MW; 33297 CN;

Initial Score = 9 Optimized Score = 9 Significance = 4.58
 Residue Identity = 33% Matches = 9 Mismatches = 18
 Gaps = 0 Conservative Substitutions = 0

X 10 20 X
 KRVDLFLTGTPDEYVEQVAQYKALPV
 || | || ||
 FPIPLPFCWLCRTLIKRIQAVVPKGVLLKAVAQVCHVPLPVGGICQCLAERYIVICLNMLLDRTLPQLVCG
 10 X 20 30 40 X 50 60 70
 LVLRCSS

5. US-08-300-510-1 (1-27)
 HYPC_ECOLI HYDROGENASE ISOENZYMES FORMATION PROTEIN HYPC.

ID HYPC_ECOLI STANDARD; PRT; 90 AA.
 AC P24191;
 DT 01-MAR-1992 (REL. 21, CREATED)
 DT 01-MAR-1992 (REL. 21, LAST SEQUENCE UPDATE)
 DT 01-JUL-1993 (REL. 26, LAST ANNOTATION UPDATE)
 DE HYDROGENASE ISOENZYMES FORMATION PROTEIN HYPC.
 GN HYPC.
 OS ESCHERICHIA COLI.
 OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; FACULTATIVELY ANAEROBIC RODS;
 OC ENTEROBACTERIACEAE.
 RN [1]
 RP SEQUENCE FROM N.A.
 RM 91194542
 RA LUTZ S., JACOBI A., SCHLENSOG V., BOEHM R., SAWERS G., BOECK A.;
 RL MOL. MICROBIOL. 5:123-135(1991).
 CC -!- FUNCTION: IS REQUIRED FOR THE FORMATION OF ALL THREE HYDROGENASE
 CC ISOENZYMES.

CC -!- SIMILARITY: BELONGS TO THE HUFF/HYPC FAMILY.
 DR EMBL; X54543; ECHYP.
 DR PIR; S15199; S15199.
 DR ECO2DBASE; A008.0; 6TH EDITION.
 DR ECGENE; EG10485; HYPC.
 SQ SEQUENCE 90 AA; 9732 MW; 39420 CN;

Initial Score = 9 Optimized Score = 12 Significance = 4.58
 Residue Identity = 41% Matches = 14 Mismatches = 13
 Gaps = 7 Conservative Substitutions = 0

```

          X      10      20      X
          KRVDVFLTLGTPDE----YVEQ---VAQYKALPV
          ||||| | | | | | | | | | |
MCIGVPGQIRITIDGNGAKVDVCGIQRDVLTLVGSCDENGQPRVGQWVLVHVGFAMSVINEAEARDTLDALQ
      10      20      X      30      40      50      60      70

NMFDVEPDVGALLYGEEK
      80      90

```

6. US-08-300-510-1 (1-27)
 KAC6_RABIT IG KAPPA CHAIN B5 VARIANT C REGION.

ID KAC6_RABIT STANDARD; PRT: 104 AA.
 AC P03984;
 DT 23-OCT-1986 (REL. 02, CREATED)
 DT 23-OCT-1986 (REL. 02, LAST SEQUENCE UPDATE)
 DT 01-APR-1988 (REL. 07, LAST ANNOTATION UPDATE)
 DE IG KAPPA CHAIN B5 VARIANT C REGION.
 OS DRYCTOLAGUS CUNICULUS (RABBIT).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
 OC EUTHERIA; LAGOMORPHA.
 RN [1]
 RP CLONE PKB5-F2, SEQUENCE FROM N.A.
 RM 84041515
 RA BERNSTEIN K.E., SKURLA R.M. JR., MAGE R.G.;
 RL NUCLEIC ACIDS RES. 11:7205-7214(1983).
 CC -!- THE CDNA FROM WHICH THIS SEQUENCE WAS DERIVED CONTAINS A
 CC TERMINATOR CODON WITHIN THE V-REGION CODING REGION. THE ORIGIN
 CC OF THIS CODON AND OF THE DIFFERENCES BETWEEN THIS AND OTHER
 CC SEQUENCED B5 C REGIONS ARE UNCLEAR. THE CDNA CLONE WAS MADE
 CC USING MRNA FROM TRYPANOSOME- INFECTED B5-HOMOZYGOUS RABBITS.
 DR PIR; A02124; K5RBV.
 DR PROSITE; PS00290; IG_MHC.
 KW IMMUNOGLOBULIN C REGION.
 FT NON_TER 1 1
 FT DISULFID 26 85
 FT DISULFID 104 104 INTERCHAIN (WITH A HEAVY CHAIN).
 SQ SEQUENCE 104 AA; 11079 MW; 62252 CN;

Initial Score = 9 Optimized Score = 11 Significance = 4.58
 Residue Identity = 34% Matches = 12 Mismatches = 15
 Gaps = 8 Conservative Substitutions = 0

```

          X      10
          KRVDVFLTLGT-
          | | |
ATLAPTVLIFPPSPAELATGTATIVCVANKYFPDGTVTWQVDCKPLTTGIETSKTPQNSDDCTYNLSSTLT
      10      20      30      40      50      60      70

      20      X
-----PDEYVEQVAQYKALPV
      ||| |||| |
KSDEYNSHDEYTCQVAQSGSGSPVVGFSRKNK
      80      90      X      100

```

7. US-08-300-510-1 (1-27)

FIMA_BORPE FIMA PROTEIN.

ID FIMA_BORPE STANDARD; PRT; 145 AA.
 AC P35076;
 DT 01-FEB-1994 (REL. 28, CREATED)
 DT 01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)
 DT 01-FEB-1994 (REL. 28, LAST ANNOTATION UPDATE)
 DE FIMA PROTEIN.
 GN FIMA.
 OS BORDETELLA PERTUSSIS.
 OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; AEROBIC RODS AND COCCI;
 OC ALCALIGENACEAE.
 RN [1]
 RP SEQUENCE FROM N.A.
 RM 93078620
 RA WILLEMS R.J., DER HEIDE H.G., MOOI F.R.;
 RL MOL. MICROBIOL. 6:2661-2671(1992).
 DR EMBL; X64876; BPFIMABC.
 SQ SEQUENCE 145 AA; 15134 MW; 107653 CN;

Initial Score = 9 Optimized Score = 11 Significance = 4.58
 Residue Identity = 37% Matches = 12 Mismatches = 15
 Gaps = 5 Conservative Substitutions = 0

X 10 20
 KRDVD-----LFLTGTPDEYVEQVAQY
 || || |||||
 MQLPTISRTALKDVGSTAGGTVDVKLTECPGALNGQQVGLFFESGGTVDYTSGNLFAYRADSQGVEQVPQT
 10 20 30 40 X 50 60 70

X
 KALPV
 || |
 KADNVQANLDGSAIHLGRNKGAQAAQTFLVSQTAGSSTYGATLRYLACYIRSGAGSIVAGNLRSGVGFSSVMY
 X 80 90 100 110 120 130 140

P

8. US-08-300-510-1 (1-27)

GTTR_RAT GLUTATHIONE S-TRANSFERASE YRS-YRS (EC 2.5.1.18) (C

ID GTTR_RAT STANDARD; PRT; 243 AA.
 AC P36971;
 DT 01-JUN-1994 (REL. 29, CREATED)
 DT 01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)
 DT 01-JUN-1994 (REL. 29, LAST ANNOTATION UPDATE)
 DE GLUTATHIONE S-TRANSFERASE YRS-YRS (EC 2.5.1.18) (CLASS-THETA).
 OS RATTUS NORVEGICUS (RAT).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
 OC EUTHERIA; RODENTIA.
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC TISSUE=LIVER;
 RM 92109741
 RA OGURA K., NISHIYAMA T., OKADA T., KAJITA J., NARIHATA H., WATABE T.,
 RA HIRATSUKA A., WATABE T.;
 RL BIOCHEM. BIOPHYS. RES. COMMUN. 181:1294-1300(1991).
 RN [2]
 RP SEQUENCE OF 1-25, AND CHARACTERIZATION.
 RM 90307730
 RA HIRATSUKA A., SEBATA N., KAWASHIMA K., OKUDA H., OGURA K., WATABE T.,

RA SATOH R., MATAYAMA I., TSUCHIDA S., ISHIKAWA I., SATO R.,
 RL J. BIOL. CHEM. 265:11973-11981(1990).
 CC -!- FUNCTION: CATALYZES THE INACTIVATION OF REACTIVE SULFATE ESTERS IN
 CC CARCINOGENIC ARYLMETHANOLS. HIGHEST ACTIVITY TOWARDS ETHACRYNIC
 CC ACID AND CUMENE HYDROPEROXIDE.
 CC -!- CATALYTIC ACTIVITY: RX + GLUTATHIONE = HX + R-S-G.
 CC -!- SUBUNIT: HOMODIMER.
 CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
 CC -!- TISSUE SPECIFICITY: HIGHEST VALUES FOUND IN LIVER FOLLOWED BY
 CC TESTIS, ADRENAL GLAND, KIDNEY, LUNG, BRAIN AND SKELETAL MUSCLE.
 CC -!- SIMILARITY: WITH OTHER GLUTATHIONE S-TRANSFERASES. BELONGS TO
 CC CLASS THETA.
 DR EMBL; D10026; RINGSTYRS.
 DR PIR; JS0618; JS0618.
 DR PIR; A37069; A37069.
 KW TRANSFERASE; MULTIGENE FAMILY.
 FT INIT_MET 0 0
 SQ SEQUENCE 243 AA; 27308 MW; 285713 CN;

Initial Score = 9 Optimized Score = 10 Significance = 4.58
 Residue Identity = 37% Matches = 10 Mismatches = 17
 Gaps = 0 Conservative Substitutions = 0

```

                X      10      20      X
                KRQVDLFLTGTPDEYVEQVAQYKALPV
                | |||      | || | ||
GLELYLDLLSQPSRAVYIFAKNGIPFQLRTVDLLKGQHLSEQFSQVNCCLKKVPVLKDGFSVLTSTAILIY
      10      20      30      40      50      X 60      70

LSSKYQVADHWYPADLQARAQVHEYLGWHADNIRGTFGVLLWTKVLGPLIGVQVPEEKVERNRNSMVLALQR
      80      90      100      110      120      130      140

LEDKFLRDRAF
      150
  
```

9. US-08-300-510-1 (1-27)

CRTC_RHOCA HYDROXYNEUROSPORENE DEHYDROGENASE (EC 1.-.-.-) (HY

ID CRTC_RHOCA STANDARD; PRT; 281 AA.
 AC P17058;
 DT 01-AUG-1990 (REL. 15, CREATED)
 DT 01-AUG-1990 (REL. 15, LAST SEQUENCE UPDATE)
 DT 01-MAY-1992 (REL. 22, LAST ANNOTATION UPDATE)
 DE HYDROXYNEUROSPORENE DEHYDROGENASE (EC 1.-.-.-) (HYDROXYNEUROSPORENE
 DE SYNTHASE).
 GN CRTC.
 OS RHODOBACTER CAPSULATUS (RHODOPSEUDOMONAS CAPSULATA).
 OC PROKARYOTA; GRACILICUTES; ANOXYPHOTOBACTERIA; PURPLE BACTERIA;
 OC RHODOSPIRILLACEAE.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SB1003, AND BEC404;
 RM 89J13663
 RA ARMSTRONG G.A., ALBERTI M., LEACH F., HEARST J.E.;
 RL MOL. GEN. GENET. 216:254-268(1989).
 CC -!- PATHWAY: CAROTENOID AND CHLOROPHYLL BIOSYNTHESIS.
 DR EMBL; X52291; RCCRTAK.
 DR EMBL; Z11165; RCPHSYNG.
 DR PIR; S04405; S04405.
 KW PHOTOSYNTHESIS; CHLOROPHYLL BIOSYNTHESIS; CAROTENOID BIOSYNTHESIS;
 KW OXIDOREDUCTASE.
 SQ SEQUENCE 281 AA; 31856 MW; 405094 CN;

Initial Score = 9 Optimized Score = 10 Significance = 4.58
 Residue Identity = 35% Matches = 11 Mismatches = 16

```

X 10
KRD----VDLFLTGTPDEYVEQ
  ||      || | |
MIAFIGSVFSPWYRWSGRREPQNHCCINMVTGTGDRFTMTDRGRSALRQSRDSFQVGPSKLTWTCKELVID
  10      20      30      40      50      60      70

20 X
VAQYKALPV
  |||
VDEWGALPKLGKLGKGRVVLTPRAVTGVEVRLTPDAGHTWRPFAPADVEVDLAPGHKWTGHGYFDANFGTRA
  80      90      100      110      120      130      140

LEEDFSFWTWGRFPLKDRTVCFYDATRLDRTKVALAV
  150      160      170      180

```

10. US-08-300-510-1 (1-27)

FVT1_HUMAN FOLLICULAR VARIANT TRANSLOCATION PROTEIN 1 PRECURS

```

ID FVT1_HUMAN STANDARD; PRT: 332 AA.
AC Q06136;
DT 01-JUN-1994 (REL. 29, CREATED)
DT 01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)
DT 01-OCT-1994 (REL. 30, LAST ANNOTATION UPDATE)
DE FOLLICULAR VARIANT TRANSLOCATION PROTEIN 1 PRECURSOR (FVT-1).
GN FVT1.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; PRIMATES.
RN [1]
RP SEQUENCE FROM N.A.
RM 93112945
RA RIMOKH R., GADOUX M., BERTHEAS M.-F., BERGER F., GAROSCIO M.,
RA DELEAGE G., GERMAIN D., MAGAUD J.-P.;
RL BLOOD 81:136-142(1993).
CC -!- SUBCELLULAR LOCATION: SECRETED (POTENTIAL).
CC -!- TISSUE SPECIFICITY: WEAKLY EXPRESSED IN NORMAL HEMATOPOIETIC
CC TISSUES. HIGHER EXPRESSION IN SOME T-CELL MALIGNANCIES AND PHA-
CC STIMULATED LYMPHOCYTES.
CC -!- DISEASE: INVOLVED IN A T(2;18)(P11;Q21) CHROMOSOMAL TRANSLOCATION
CC WITH A IG J KAPPA CHAIN REGION THAT PRODUCES AN ONCOGENE
CC RESPONSIBLE FOR FOLLICULAR LYMPHOMA (ALSO KNOWN AS TYPE II CHRONIC
CC LYMPHATIC LEUKEMIA).
DR EMBL: S51904; HSFVT1A.
DR PIR: S37652; S37652.
DR MIM: 136440; 11TH EDITION.
KW PROTO-ONCOGENE; CHROMOSOMAL TRANSLOCATION; SIGNAL.
FT SIGNAL 1 25 POTENTIAL.
FT CHAIN 26 332 FOLLICULAR VARIANT TRANSLOCATION
FT PROTEIN 1.
SQ SEQUENCE 332 AA; 36187 MW; 585292 CN;

```

```

Initial Score = 9 Optimized Score = 9 Significance = 4.58
Residue Identity = 33% Matches = 9 Mismatches = 18
Gaps = 0 Conservative Substitutions = 0

```

```

MLLLAAFLVAFVLLLYMVSPLISPKPLALPGAHVVTGSSGIGKCIAIECYKQGAFITLVARNEDKLLQA
  10      20      30      40      50      60      70

X 10 20 X
KRDVDLFLTGTPDEYVEQVAQYKALPV
  ||      || | |
KKEIEMHSINDKQVVLCSVDVSQDYNQVENVIKQAGEKLGPDMLVNCAGMAVSGKFEDLEVSTFERLMSI
  80      90      100      110      120      130      140

```

11. US-08-300-510-1 (1-27)

E13B_HORVU GLUCAN ENDO-1,3-BETA-GLUCOSIDASE GII PRECURSOR (EC

ID E13B_HORVU STANDARD; PRT; 334 AA.
AC P15737;
DT 01-APR-1990 (REL. 14, CREATED)
DT 01-APR-1990 (REL. 14, LAST SEQUENCE UPDATE)
DT 01-OCT-1994 (REL. 30, LAST ANNOTATION UPDATE)
DE GLUCAN ENDO-1,3-BETA-GLUCOSIDASE GII PRECURSOR (EC 3.2.1.39) ((1->3)-
DE BETA-GLUCAN ENDOHYDROLASE) ((1->3)-BETA-GLUCANASE ISOENZYME GII).
OS HORDEUM VULGARE (BARLEY).
OC EUKARYOTA; PLANTA; EMBRYOPHYTA; ANGIOSPERMAE; MONOCOTYLEDONEAE;
OC CYPERALES; GRAMINEAE.
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 29-68.
RC STRAIN=CV. CLIPPER;
RM 93357431
RA HOEJ P.B., HARTMAN D.J., MORRICE N.A., DOAN D.N.P., FINCHER G.B.;
RL PLANT MOL. BIOL. 13:31-42(1989).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. PIGGY;
RM 91107649
RA LEAH R., TOMMERUP H., SVENDSEN I., MUNDY J.;
RL J. BIOL. CHEM. 266:1564-1573(1991).
RN [3]
RP SEQUENCE OF 258-332 FROM N.A.
RC TISSUE=LEAF;
RA JUTIDAMRONGPHAN W., MACKINNON G., MANNERS J., SIMPSON R.S.,
RA SCOTT K.J.;
RL SUBMITTED (AUG-1989) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [4]
RP SEQUENCE OF 29-334.
RA BALLANCE G.M., SVENDSEN I.;
RL CARLSBERG RES. COMMUN. 53:411-419(1988).
CC -!- FUNCTION: MAY PROVIDE A DEGREE OF PROTECTION AGAINST MICROBIAL
CC INVASION OF GERMINATED BARLEY GRAIN THROUGH ITS ABILITY TO DEGRADE
CC FUNGAL CELL WALL POLYSACCHARIDES.
CC -!- CATALYTIC ACTIVITY: HYDROLYSIS OF 1,3-BETA-D-GLUCOSIDIC LINKAGES
CC IN 1,3-BETA-D-GLUCANS.
CC -!- SIMILARITY: BELONGS TO FAMILY 17 OF GLYCOSYL HYDROLASES.
DR EMBL; X15205; HV13BGE.
DR EMBL; M62907; HVCBGL32.
DR EMBL; X16274; HVB13GLU.
DR EMBL; M23548; HVGEH.
DR PIR; S05510; S05510.
DR PIR; A31800; A31800.
DR PROSITE; PS00587; GLYCOSYL_HYDROL_F17.
KW HYDROLASE; GLUCOSIDASE; SIGNAL; MULTIGENE FAMILY.
FT SIGNAL 1 28
FT CHAIN 29 334 GLUCAN ENDO-1,3-BETA-GLUCOSIDASE GII.
FT ACT_SITE 246 246 POTENTIAL.
FT ACT_SITE 259 259 POTENTIAL.
FT CONFLICT 12 12 A -> V (IN REF. 2).
FT CONFLICT 71 71 L -> V (IN REF. 2).
SQ SEQUENCE 334 AA; 35193 MW; 549197 CN;

Initial Score = 9 Optimized Score = 9 Significance = 4.58
Residue Identity = 33% Matches = 9 Mismatches = 18
Gaps = 0 Conservative Substitutions = 0

NEVGGATQSLFAMRNENALSAAGLGATVSTSIKFDVANSFPFSAQVFNATVTVARLLASTGAFLL
 130 140 150 160 170 180 190
 X 10 20 X
 KRVDLFLTGTPDEYVEQVAQYKALPV
 || || | || ||
 ANVYPYFAYRDNPGSISLNYATFQPGTTVRDQNNGLTYTSLFDAMVDAVYAALKAGAPAVKVVVSESGWPS
 200 210 220 230 240 250 260
 AGGFAASAGNARTYNQGLINHVGGGTPKKREALETYIFAMFNENQKTGDATERSFGLFNPDKSPAYNIQF
 270 280 290 300 310 320 330

12. US-08-300-510-1 (1-27)

ALGP_PSEAE TRANSCRIPTIONAL REGULATORY PROTEIN ALGP (ALGINATE

ID ALGP_PSEAE STANDARD; PRT; 340 AA.
 AC P15276;
 DT 01-APR-1990 (REL. 14, CREATED)
 DT 01-APR-1990 (REL. 14, LAST SEQUENCE UPDATE)
 DT 01-OCT-1993 (REL. 27, LAST ANNOTATION UPDATE)
 DE TRANSCRIPTIONAL REGULATORY PROTEIN ALGP (ALGINATE REGULATORY PROTEIN
 DE ALGR3).
 GN ALGP OR ALGR3.
 OS PSEUDOMONAS AERUGINOSA.
 OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; AEROBIC RODS AND COCCI;
 OC PSEUDOMONADACEAE.
 RN [1]
 RP SEQUENCE FROM N.A.
 RM 90108714
 RA KATO J., CHU L., KITANO K., DEVAULT J.D., KIMBARA K.,
 RA CHAKRABARTY A.M., MISRA T.K.;
 RL GENE 84:31-38(1989).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=8882;
 RM 90222135
 RA KATO J., MISRA T.K., CHAKRABARTY A.M.;
 RL PROC. NATL. ACAD. SCI. U.S.A. 87:2887-2891(1990).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=8830;
 RM 91008921
 RA DERETIC V., KONYECSNI W.M.;
 RL J. BACTERIOL. 172:5544-5554(1990).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=PAD / PA02003;
 RM 90236911
 RA KONYECSNI W.M., DERETIC V.;
 RL J. BACTERIOL. 172:2511-2520(1990).
 CC -!- FUNCTION: THE PROMOTER FOR A CRITICAL ALGINATE BIOSYNTHETIC
 CC GENE, ALGD, ENCODING GDP-MANNOSE DEHYDROGENASE, IS ACTIVATED ONLY
 CC UNDER CONDITIONS REMINISCENT OF THE CYSTIC FIBROSIS LUNG (I.E.,
 CC UNDER HIGH OSMOLARITY), AND AT LEAST TWO REGULATORY GENES, ALGP
 CC AND ALGQ, HAVE BEEN IMPLICATED IN THIS ACTIVATION PROCESS.
 CC -!- DISEASE: ALGINATE IS AN EXOPOLYSACCHARIDE PRODUCED BY STRAINS
 CC OF P. AERUGINOSA DURING INFECTION IN THE RESPIRATORY TRACT OF
 CC CYSTIC FIBROSIS PATIENTS.
 CC -!- CARBOXY-TERMINAL BINDS TO DNA. IT IS UNKNOWN WHETHER BINDING IS
 CC SPECIFIC OR NON-SPECIFIC.
 CC -!- SIMILARITY: TO EUKARYOTIC HISTONES H1.
 DR EMBL; M30145; PAARGRA.
 DR EMBL; M57551; PAALGP.
 DR EMBL; M32077; PAALALPO.
 DR EMBL; M35259; PAALGR3A.

DR PIR: J001467 J00148.
 DR PIR: A35630; A35630.
 DR PIR: A36128; A36128.
 KW ALGINATE BIOSYNTHESIS; TRANSCRIPTION REGULATION; ACTIVATOR;
 KW DNA-BINDING; REPEAT.
 FT CONFLICT 28 28 G -> D (IN REF. 4).
 FT CONFLICT 157 158 NA -> KR (IN REF. 3).
 FT CONFLICT 157 158 NA -> KP (IN REF. 4).
 FT CONFLICT 173 173 K -> KPAK (IN REF. 4).
 FT CONFLICT 177 177 A -> G (IN REF. 3 AND 4).
 FT CONFLICT 184 184 A -> T (IN REF. 4).
 FT CONFLICT 219 220 NA -> KP (IN REF. 3 AND 4).
 FT CONFLICT 242 242 A -> T (IN REF. 4).
 FT CONFLICT 262 263 HV -> PA (IN REF. 4).
 FT CONFLICT 268 268 A -> AKPVAKSAA (IN REF. 4).
 FT CONFLICT 278 279 NA -> KP (IN REF. 3 AND 4).
 FT CONFLICT 299 299 A -> T (IN REF. 4).
 FT CONFLICT 308 309 NA -> KP (IN REF. 3 AND 4).
 FT CONFLICT 316 316 P -> T (IN REF. 2).
 SQ SEQUENCE 340 AA; 33187 MW; 440087 CN;

Initial Score = 9 Optimized Score = 9 Significance = 4.58
 Residue Identity = 33% Matches = 9 Mismatches = 18
 Gaps = 0 Conservative Substitutions = 0

KPVTTPLHLLQQLSHSLVEHLEGACKQALVDSEKLLAKLEKQRGKAGEKLHKARTKLQDAAKAGKTKAQAKA
 10 20 30 40 50 60 70

X 10 20 X
 KRVDLFLTGTPDEYVEQVAQYKALPV

RETISDLEEALDTLKARQADTRTYIVGLKRDVQESLKLQGVGVKEAAGKALESRKAKPATKPAAKAAAKP
 80 90 100 X 110 120 130 X 140

AVKTVAAANAAKPAKPAKPAAKTAAKPAKPAKPAKPAKPAKTAATAKPAKPAKPAKPAKPAANAA
 150 160 170 180 190 200 210 220

AKTAAKPAK
 230

13. US-08-300-510-1 (1-27) CHLI_EUGGR PROBABLE MAGNESIUM-CHELATASE SUBUNIT.

ID CHLI_EUGGR STANDARD; PRT: 348 AA.
 AC P31205;
 DT 01-JUL-1993 (REL. 26, CREATED)
 DT 01-JUL-1993 (REL. 26, LAST SEQUENCE UPDATE)
 DT 01-JUL-1993 (REL. 26, LAST ANNOTATION UPDATE)
 DE PROBABLE MAGNESIUM-CHELATASE SUBUNIT.
 GN CHLI OR CCSA.
 OS EUGLENA GRACILIS.
 OG CHLOROPLAST.
 OC EUKARYOTA; PLANTA; PHYCOPHYTA; EUGLENOPHYTA.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Z;
 RM 92299087
 RA ORSAT B., MONFORT A., CHATELLARD P., STUTZ E.;
 RL FEBS LETT. 303:181-184(1992).
 CC -!- FUNCTION: INVOLVED IN CHLOROPLAST PIGMENT BIOSYNTHESIS (PROBABLE).
 CC -!- PATHWAY: CHLOROPHYLL BIOSYNTHESIS.
 CC -!- SIMILARITY: TO R.CAPSULATUS BCHI AND A.THALIANA CHLI.
 DR EMBL: Z11874; CHEGZ.
 DR EMBL: X65484; EGCCSA.
 DR PIR: S21383; S21383.

UN PIR: 554474; 554474.
 KW PHOTOSYNTHESIS; CHLOROPHYLL BIOSYNTHESIS; CHLOROPLAST.
 SQ SEQUENCE 348 AA; 39307 MW; 615982 CN;

Initial Score = 9 Optimized Score = 9 Significance = 4.58
 Residue Identity = 33% Matches = 9 Mismatches = 18
 Gaps = 0 Conservative Substitutions = 0

MNKKTNERPVFPFTSIVGQEEMLSLILNVIDPKIGGVIMGDRGTGKSTIVRALVDLLPPIDVIENDPYNS
 10 20 30 40 50 60 70
 X 10 20 X
 KRVDLFLTGTPDEYVEQVAQYKALPV
 ||| ||| ||
 DPYDTELMSDDVLEKIKKNEKVSIIQVKTPMVDLPLGGTEDRVCGTIDIEKAISEGKKAFEPGLLAGANRGI
 80 90 100 110 120 130 140
 LYVDEVNLLDDHLVDVLLDSAASQWNTVEREGVSICHPARFILVGSGNPEEGELRPQLLDRFGMHAQIKTLK
 150 160 170 180 190 200 210
 EPALRVKIVQQ
 220

14. US-08-300-510-1 (1-27)
 PRTZ_HORVU PROTEIN Z (Z4) (MAJOR ENDOSPERM ALBUMIN).

ID PRTZ_HORVU STANDARD; PRT; 399 AA.
 AC P06293;
 DT 01-JAN-1988 (REL. 06, CREATED)
 DT 01-NOV-1991 (REL. 20, LAST SEQUENCE UPDATE)
 DT 01-MAR-1992 (REL. 21, LAST ANNOTATION UPDATE)
 DE PROTEIN Z (Z4) (MAJOR ENDOSPERM ALBUMIN).
 GN PAZ1.
 OS HORDEUM VULGARE (BARLEY).
 OC EUKARYOTA; PLANTA; EMBRYOPHYTA; ANGIOSPERMAE; MONOCOTYLEDONEAE;
 OC CYPERALES; GRAMINEAE.
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC STRAIN=CV. CARLSBERG II; TISSUE=GRAIN;
 RM 91099324
 RA BRANDT A., SVENDSEN I., HEJGAARD J.;
 RL EUR. J. BIOCHEM. 194:499-505(1990).
 RN [2]
 RP SEQUENCE OF 220-399 FROM N.A., AND PARTIAL SEQUENCE.
 RC STRAIN=CV. CARLSBERG II; TISSUE=GRAIN;
 RA HEJGAARD J., RASMUSSEN S.K., BRANDT A., SVENDSEN I.;
 RL FEBS LETT. 180:89-94(1985).
 CC -!- FUNCTION: A MAJOR COMPONENT OF THE ENDOSPERM ALBUMIN, THIS PROTEIN
 CC ACTS AS A STORAGE PROTEIN DURING GRAIN FILLING, CONTRIBUTING A
 CC SUBSTANTIAL PART OF THE GRAIN'S LYSINE.
 CC -!- TISSUE SPECIFICITY: IS ACCUMULATED AND STORED IN THE ENDOSPERM,
 CC WHERE IT EXISTS IN A FREE AND A BOUND FORM.
 CC -!- DEVELOPMENTAL STAGE: SYNTHESIZED 10-25 DAYS AFTER FERTILIZATION
 CC (DEVELOPING ENDOSPERM).
 CC -!- INDUCTION: ITS EXPRESSION IS REGULATED BY THE "HIGH LYSINE"
 CC ALLELES LYS1 AND LYS3A.
 CC -!- SIMILARITY: WITH SERPINS. THIS SUGGESTS THAT THIS PROTEIN ALSO HAS
 CC AN INHIBITORY FUNCTION DURING FILLING OR GERMINATION.
 CC -!- THERE SEEM TO BE TWO Z PROTEINS: Z4 (FROM CHROMOSOME 4) AND Z7
 CC (FROM CHROMOSOME 7).
 DR EMBL; X51726; HVPAZ1.
 DR EMBL; X05902; HVPROTZ.
 DR PIR; A01252; DXBHZ.
 DR PIR; S13822; S13822.
 DR PROSITE; PS00284; SERPIN.

RW SERPIN7 STORAGE PROTEIN; MULTIGENE FAMILY.
 FT DOMAIN 36 56 SIGNAL FOR TARGETING PROTEIN Z4 INTO
 FT THE ER LUMEN (POTENTIAL).
 FT ACT_SITE 357 357 REACTIVE BOND (POTENTIAL).
 SQ SEQUENCE 399 AA; 43276 MW; 857901 CN;

Initial Score = 9 Optimized Score = 9 Significance = 4.58
 Residue Identity = 33% Matches = 9 Mismatches = 18
 Gaps = 0 Conservative Substitutions = 0

ATDVRLSIAHQTRFALRLRSAISSNPERRAAGNVAFSPLSLHVALSLITAGAAATRDQLVAILGDGGAGDAKE
 10 20 30 40 50 60 70
 X 10 20 X
 KRVDVFLFTGTPDEYVEQVAGYKALPV
 || | | |
 LNALAEQVVQFVLANESSTGGPRIAFANGIFVDASLSLKPSFEELAVCCQYKAKTQSVDFQHKLEAVGQVNS
 80 90 100 X 110 120 130 X 140
 WVEQVTTGLIKQILPPGSVDNTTKLILGNALYFKGAWDQKFDESNTKCDSFHLLDGSSIQTQFMSSTKKQYI
 150 160 170 180 190 200 210 220
 SSSDNLKV LKL
 230

15. US-08-300-510-1 (1-27)
 PRE2_STAAU PLASMID RECOMBINATION ENZYME (MOBILIZATION PROTEIN)

ID PRE2_STAAU STANDARD; PRT; 420 AA.
 AC P22490;
 DT 01-AUG-1991 (REL. 19, CREATED)
 DT 01-AUG-1991 (REL. 19, LAST SEQUENCE UPDATE)
 DT 01-JUN-1994 (REL. 29, LAST ANNOTATION UPDATE)
 DE PLASMID RECOMBINATION ENZYME (MOBILIZATION PROTEIN).
 GN PRE OR MOB.
 OS STAPHYLOCOCCUS AUREUS.
 OG PLASMID PUB110.
 OC PROKARYOTA; FIRMICUTES; COCCI; MICROCOCCACEAE.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA BASHKIROV V.I., MIL'SHINA N.V., PROZOROV A.A.;
 RL GENETIKA 22:823-831(1986).
 CC -!- FUNCTION: THE INTERACTION OF THE RSA SITE AND THE PRE PROTEIN
 CC MAY NOT ONLY SERVES A FUNCTION IN PLASMID MAINTENANCE, BUT ALSO
 CC MAY CONTRIBUTES TO THE DISTRIBUTION OF SMALL ANTIBIOTIC RESISTANCE
 CC PLASMIDS AMONG GRAM-POSITIVE BACTERIA.
 CC -!- SIMILARITY: TO OTHER PRE PROTEINS (FROM PLASMIDS PUB110, PMV158,
 CC PE194, PT181, PTB913), IN THEIR N-TERMINAL ONLY.
 CC -!- PRE PROTEINS CONTAIN CONSERVED POSITIVELY CHARGED AMINO ACIDS
 CC PROBABLY INVOLVED IN THE BINDING OF THE PRE PROTEIN TO THE RSA
 CC SITE.
 DR EMBL; M37273; PPKANRCG.
 KW PLASMID; DNA-BINDING.
 FT BINDING 44 44 DNA (POTENTIAL).
 FT BINDING 114 114 DNA (POTENTIAL).
 SQ SEQUENCE 420 AA; 49660 MW; 900437 CN;

Initial Score = 9 Optimized Score = 9 Significance = 4.58
 Residue Identity = 33% Matches = 9 Mismatches = 18
 Gaps = 0 Conservative Substitutions = 0

MFGLQKEIMKTEKKPTKNVVISERDYKNLVTAARDNDR LKQHVRNLHSTDHAREYKKLSKEHGQVKEKYSGL
 250 260 270 280 290 300 310
 X 10 20 X

RRDVLFLTGIPDEYVEQVAVYKALFV
 IIII I I I II
 VERFNEVNDYNELLEENKSLKSKISDLKRDVSLIYESTKEFLKERTDGLKAFKNVFKGFVDKVKDKTAQFQ
 320 330 340 X 350 360 370 380

EKHDLEPKKNEFELTHNREVKKERSRDQGMSL
 390 400 410 420

> O <
 OI IO IntelliGenetics
 > O <

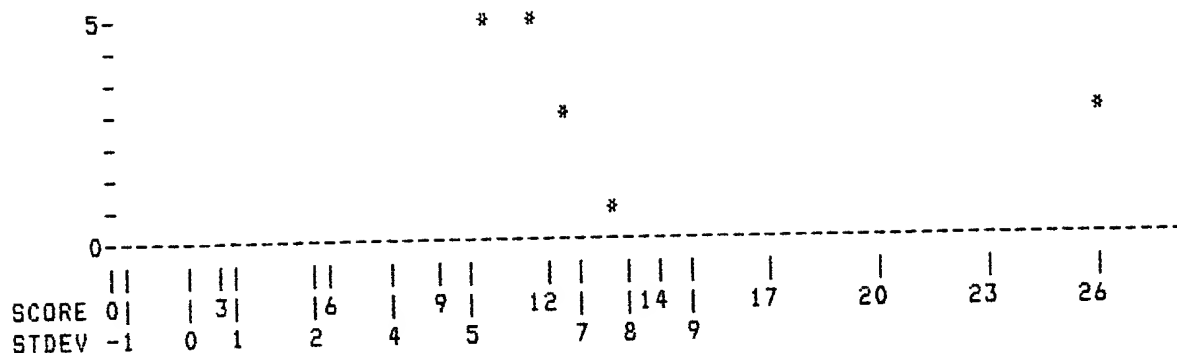
FastDB - Fast Pairwise Comparison of Sequences
 Release 5.4

Results file 2-pir.res made by on Fri 24 Mar 95 7:48:03-PST.

Query sequence being compared: US-08-300-510-2 (1-27)
 Number of sequences searched: 75511
 Number of scores above cutoff: 4500

Results of the initial comparison of US-08-300-510-2 (1-27) with:
 Data bank : PIR 43, all entries

100000-
 -
 N -
 U50000-
 M -
 B -
 E - *
 R - * *
 -
 O - *
 F10000-
 -
 S -
 E 5000-
 Q - *
 U -
 E *
 N -
 C -
 E - *
 S 1000-
 -
 500-
 -
 -
 -
 -
 -
 100- *
 -
 -
 50- *
 -
 -
 -
 -
 10-
 -



PARAMETERS

Similarity matrix	Unitary	K-tuple	2
Mismatch penalty	1	Joining penalty	20
Gap penalty	1.00	Window size	27
Gap size penalty	0.05		
Cutoff score	0		
Randomization group	0		
Initial scores to save	40	Alignments to save	15
Optimized scores to save	0	Display context	100

SEARCH STATISTICS

Scores:	Mean	Median	Standard Deviation
	3	4	1.42

Times:	CPU	Total Elapsed
	00:00:59.05	00:01:00.00

Number of residues:	22468834
Number of sequences searched:	75511
Number of scores above cutoff:	4500

Cut-off raised to 3.
 Cut-off raised to 4.
 Cut-off raised to 5.
 Cut-off raised to 6.
 Cut-off raised to 7.

The scores below are sorted by initial score.
 Significance is calculated based on initial score.

A 100% identical sequence to the query sequence was not found.

The list of best scores is:

Sequence Name	Description	Length	Init. Score	Opt. Score	Sig.	Frame
	**** 16 standard deviations above mean ****					
1. JC1126	major allergen chain 1 precur	88	26	26	16.17	0
2. JC1136	major allergen chain 1 precur	92	26	26	16.17	0
	**** 7 standard deviations above mean ****					
3. GNNY2F	genome polyprotein - foot-and	2333	13	14	7.03	0
	**** 6 standard deviations above mean ****					
4. A53283	major cat allergen Fel d I al	40	12	12	6.33	0
5. S37077	genome polyprotein - foot-and	2336	12	13	6.33	0
	**** 5 standard deviations above mean ****					
6. S40064	3-deoxy-D-manno-2-octulosonic	411	11	11	5.62	0
7. S28562	3-deoxy-D-manno-2-octulosonic	411	11	11	5.62	0

8. GNO431	RNA-directed RNA polymerase (470	11	12	5.62	0
9. GNNYF	genome polyprotein - foot-and	2332	11	12	5.62	0
	**** 4 standard deviations above mean ****					
10. S02068	RNA-directed RNA polymerase (470	10	11	4.92	0
11. S10340	DNA-directed RNA polymerase (982	10	10	4.92	0
12. S00964	hypothetical protein 6 - yeas	982	10	10	4.92	0
13. GNNY4F	genome polyprotein - foot-and	2332	10	11	4.92	0
14. C30305	submandibular gland protein (91	9	10	4.22	0
15. A35072	nonhistone chromosomal protei	93	9	9	4.22	0
16. B35072	nonhistone chromosomal protei	99	9	9	4.22	0
17. HRTHM	myohemerythrin - sipunculid (118	9	11	4.22	0
18. S45108	hypothetical protein 2 - Erwi	151	9	9	4.22	0
19. S29037	Na+-transporting ATP synthase	163	9	11	4.22	0
20. S12620	Na+-transporting ATP synthase	163	9	11	4.22	0
21. S24369	Na+-transporting ATP synthase	163	9	11	4.22	0
22. CFXCA	C-phycoerythrin alpha chain -	164	9	10	4.22	0
23. S23323	Na+-transporting ATP synthase	168	9	11	4.22	0
24. S15274	cutR protein - Streptomyces l	217	9	9	4.22	0
25. S44182	Phl p I allergen - Common tim	263	9	9	4.22	0
26. S24232	npl protein - Listeria monocy	271	9	10	4.22	0
27. S46030	hypothetical membrane protein	347	9	9	4.22	0
28. KIUTGC	phosphoglycerate kinase (EC 2	421	9	11	4.22	0
29. S19722	dihydrolipoamide acetyltransf	430	9	9	4.22	0
30. S42206	enolase (EC 4.2.1.11) - Plasm	446	9	9	4.22	0
31. DEBY4	alcohol dehydrogenase (EC 1.1	465	9	9	4.22	0
32. S42306	gene 4B protein - phage T7	503	9	10	4.22	0
33. A60280	bacillolysin homolog (EC 3.4.	510	9	10	4.22	0
34. A24031	genome polyprotein - foot-and	529	9	10	4.22	0
35. VHIVC8	nucleoprotein - influenza C v	565	9	9	4.22	0
36. S42304	gene 4A protein - phage T7	566	9	10	4.22	0
37. S07508	DNA primase - phage T3	566	9	10	4.22	0
38. YDBPA7	DNA primase chain A - phage T	566	9	10	4.22	0
39. C25035	colicin Ia - Escherichia coli	626	9	9	4.22	0
40. D25035	colicin Ib - Escherichia coli	626	9	9	4.22	0

1. US-08-300-510-2 (1-27)

JC1126 major allergen chain 1 precursor B - cat

```

ENTRY      JC1126      #type complete
TITLE      major allergen chain 1 precursor B - cat
ORGANISM    #formal_name Felis silvestris catus #common_name domestic cat
DATE        31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change
            31-Dec-1993
ACCESSIONS  JC1126
REFERENCE   JC1126
            #authors    Griffith, I.J.; Craig, S.; Pollock, J.; Yu, X.B.;
                        Morgenstern, J.P.; Rogers, B.L.
            #journal     Gene (1992) 113:263-268
            #title       Expression and genomic structure of the genes encoding FdI,
                        the major allergen from the domestic cat.
            #accession   JC1126
            ##molecule_type DNA
            ##residues    1-88 ##label GRI
GENETICS
            #gene        Ch1
            #introns      17/1; 79/3
FEATURE
            1-18         #domain signal sequence #status predicted #label SIG\
            19-88         #product major allergen chain 1 #status predicted #label
                        MAT
SUMMARY
            #length 88   #molecular-weight 9586   #checksum 4095
SEQUENCE

```

```

Initial Score      =      26   Optimized Score      =      26   Significance = 16.17
Residue Identity   =     96%   Matches               =      26   Mismatches   =      1

```

```

X      10      20
KALPVVLENARILKNCVDAKMTEEDK
|||||
MLDAALPPCPTVAATADCEICPAVKRDVDLFLTGTPDEYVEQVAQYNALPVVLENARILKNCVDAKMTEEDK
10      20      30      40      X 50      60      70

```

```

X
E
|
ENALSVLDKIYTSPLC
X      80

```

2. US-08-300-510-2 (1-27)

JC1136 major allergen chain 1 precursor A - cat

```

ENTRY      JC1136      #type complete
TITLE      major allergen chain 1 precursor A - cat
ORGANISM    #formal_name Felis silvestris catus #common_name domestic cat
DATE        31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change
            31-Dec-1993
ACCESSIONS  JC1136
REFERENCE   JC1126
#authors    Griffith, I.J.; Craig, S.; Pollock, J.; Yu, X.B.;
            Morgenstern, J.P.; Rogers, B.L.
#journal     Gene (1992) 113:263-268
#title       Expression and genomic structure of the genes encoding FdI,
            the major allergen from the domestic cat.
#accession   JC1136
##molecule_type DNA
##residues   1-92 ##label GRI
GENETICS
#gene        Ch1
#introns     21/1; 83/3
FEATURE
1-22         #domain signal sequence #status predicted #label SIG\
23-92        #product major allergen chain 1 #status predicted #label
            MAT
SUMMARY      #length 92 #molecular-weight 10072 #checksum 4988
SEQUENCE

```

```

Initial Score = 26 Optimized Score = 26 Significance = 16.17
Residue Identity = 96% Matches = 26 Mismatches = 1
Gaps = 0 Conservative Substitutions = 0

```

```

X      10      20
KALPVVLENARILKNCVDAKMT
|||||
MKGACVLVLLWAALLLISGGNCEICPAVKRDVDLFLTGTPDEYVEQVAQYNALPVVLENARILKNCVDAKMT
10      20      30      40      50      60      70

```

```

X
EEDKE
|||||
EEDKENALSVLDKIYTSPLC
X 80      90

```

3. US-08-300-510-2 (1-27)

GNNY2F genome polyprotein - foot-and-mouth disease virus

```

ENTRY      GNNY2F      #type complete
TITLE      genome polyprotein - foot-and-mouth disease virus A (strain
            A[10161])

```

CONTAINS coat protein VP1; coat protein VP2; coat protein VP3; coat
 protein VP4; core protein p52; genome-linked protein VPg1;
 genome-linked protein VPg2; genome-linked protein VPg3;
 nonstructural protein p20a; nonstructural protein p20b;
 RNA-directed RNA polymerase (EC 2.7.7.48)

ORGANISM #formal_name Aphthovirus A #common_name foot-and-mouth
 disease virus A

DATE 17-Dec-1982 #sequence_revision 28-Aug-1985 #text_change
 31-Dec-1993

ACCESSIONS A93508; A91491; S30753

REFERENCE A93508
 #authors Carroll, A.R.; Rowlands, D.J.; Clarke, B.E.
 #journal Nucleic Acids Res. (1984) 12:2461-2472
 #title The complete nucleotide sequence of the RNA coding for the
 primary translation product of foot and mouth disease
 virus.
 #cross-references MUID:84169547
 #accession A93508
 ##molecule_type genomic RNA
 ##residues 1-2333 ##label CAR
 ##cross-references GB:X00429

REFERENCE A91491
 #authors Boothroyd, J.C.; Harris, T.J.R.; Rowlands, D.J.; Lowe, P.A.
 #journal Gene (1982) 17:153-161
 #title The nucleotide sequence of cDNA coding for the structural
 proteins of foot-and-mouth disease virus.
 #cross-references MUID:82211814
 #accession A91491
 ##molecule_type genomic RNA
 ##residues 115-395,'C',397-631,'L',633-1048 ##label B00
 ##cross-references GB:V01130

REFERENCE S30753
 #authors Sangar, D.V.; Newton, S.E.; Rowlands, D.J.; Clarke, B.E.
 #journal Nucleic Acids Res. (1987) 15:3305-3315
 #title All foot and mouth disease virus serotypes initiate protein
 synthesis at two separate AUGs.
 #accession S30753
 ##molecule_type genomic RNA
 ##residues 1-32 ##label SAN
 ##cross-references EMBL:M31575

CLASSIFICATION #superfamily foot-and-mouth disease virus genome polyprotein

KEYWORDS coat protein; core protein; genome-linked protein;
 nonstructural protein; nucleotidyltransferase; polyprotein

FEATURE
 1-204 #product nonstructural protein p20a #label NPA\
 205-286 #product coat protein VP4 #label VP4\
 287-504 #product coat protein VP2 #label VP2\
 505-725 #product coat protein VP3 #label VP3\
 726-937 #product coat protein VP1 #label VP1\
 938-1578 #product core protein p52 #label CPP\
 1579-1601 #product genome-linked protein VPg1 #label GL1\
 1602-1625 #product genome-linked protein VPg2 #label GL2\
 1626-1649 #product genome-linked protein VPg3 #label GL3\
 1650-1863 #product nonstructural protein p20b #label NPB\
 1864-2333 #product RNA-directed RNA polymerase #label RRP

SUMMARY #length 2333 #molecular-weight 259646 #checksum 7155

SEQUENCE

Initial Score = 13 Optimized Score = 14 Significance = 7.03
 Residue Identity = 41% Matches = 15 Mismatches = 12
 Gaps = 9 Conservative Substitutions = 0

GLFAYKAATRAGYCGGAVLAKDGADTFIVGTHSAGGNGVGYCSCVSRSHLQKMKAHVDPEPHHEGLIVDTRD
 1800 1810 1820 1830 1840 1850 1860 1870

RAL-----PVVLENARILKNCVDAKMTTEEDKE
 || ||| | |||||
 VEERVHVMRKTKLAPTVA YGVFNPEFGPAALSNKDPRLNEGVLDDVIFSKHKGD AKMTEEDKALFRRCAAD
 1880 1890 1900 1910 1920 1930 X 1940
 YASRLHSVLGTANAPLSIYEAIKGV DGLDAMEPDTAPGLPWALQGKRRGALIDFENGTVGPEVEAALKLMEK
 1950 1960 1970 1980 1990 2000 2010
 REYKFACGTFLKDEIRPMEK
 2020 2030

4. US-08-300-510-2 (1-27)
A53283 major cat allergen Fel d I alpha chain - cat (frag

ENTRY A53283 #type fragment
 TITLE major cat allergen Fel d I alpha chain - cat (fragment)
 ORGANISM #formal_name Felis silvestris catus #common_name domestic cat
 DATE 12-May-1994 #sequence_revision 12-May-1994 #text_change 12-May-1994
 ACCESSIONS A53283
 REFERENCE A53283
 #authors Duffort, D.A.; Carreira, J.; Nitti, G.; Polo, F.; Lombardero, M.
 #journal Mol. Immunol. (1991) 28:301-309
 #title Studies on the biochemical structure of the major cat allergen Felis domesticus I.
 #accession A53283
 ##status preliminary
 ##molecule_type protein
 ##residues 1-40 ##label DUF
 SUMMARY #length 40 #checksum 3032
 SEQUENCE

Initial Score	=	12	Optimized Score	=	12	Significance	=	6.33
Residue Identity	=	100%	Matches	=	12	Mismatches	=	0
Gaps	=	0	Conservative Substitutions	=			=	0

X 10 X 20
 KALPVVLENARILKNCVDAKMTTEEDKE
 |||||
 EICPAVKRDVDLFLTGTPDEYVEQVAQYKALPVVLENARI
 10 20 30 40

5. US-08-300-510-2 (1-27)
S37077 genome polyprotein - foot-and-mouth disease virus

ENTRY S37077 #type complete
 TITLE genome polyprotein - foot-and-mouth disease virus A (strain A22/550 Azerbaijan 65)
 CONTAINS coat protein VP1; coat protein VP2; coat protein VP3; coat protein VP4; core protein p14; core protein p19; core protein p41; core protein X; genome-linked protein VPg1; genome-linked protein VPg2; genome-linked protein VPg3; nonstructural protein p20a; proteinase (EC 3.4.-.-); RNA-directed RNA polymerase (EC 2.7.7.48)
 ORGANISM #formal_name Aphthovirus A #common_name foot-and-mouth disease virus A
 DATE 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 31-Dec-1993
 ACCESSIONS S37077
 REFERENCE S37077
 #authors Sosnovtsev, S.V.; Onischenko, A.M.; Petrov, N.A.; Kalashnikova, T.I.; Mamaeva, N.V.; Drygin, V.Y.; Perevozchikova, N.A.; Vasilenko, S.K.

#submission submitted to the EMBL data library, August 1993
 #accession S37077
 ##molecule_type genomic RNA
 ##residues 1-2336 ##label S05
 ##cross-references EMBL:X74812
 CLASSIFICATION #superfamily foot-and-mouth disease virus genome polyprotein
 KEYWORDS coat protein; core protein; genome-linked protein;
 nonstructural protein; nucleotidyltransferase; polyprotein

FEATURE
 1-217 #product nonstructural protein p20a #status predicted
 #label NPA\
 218-286 #product coat protein VP4 #status predicted #label VP4\
 287-504 #product coat protein VP2 #status predicted #label VP2\
 505-724 #product coat protein VP3 #status predicted #label VP3\
 725-938 #product coat protein VP1 #status predicted #label VP1\
 939-954 #product core protein X #status predicted #label CPX\
 955-1108 #product core protein p14 #status predicted #label C14\
 1109-1426 #product core protein p41 #status predicted #label C41\
 1427-1579 #product core protein p19 #status predicted #label C19\
 1580-1602 #product genome-linked protein VPg1 #status predicted
 #label VG1\
 1603-1626 #product genome-linked protein VPg2 #status predicted
 #label VG2\
 1627-1650 #product genome-linked protein VPg3 #status predicted
 #label VG3\
 1651-1863 #product proteinase #status predicted #label PTS\
 1864-2333 #product RNA-directed RNA polymerase #status predicted
 #label RRP

SUMMARY #length 2336 #molecular-weight 259983 #checksum 4399
 SEQUENCE

Initial Score = 12 Optimized Score = 13 Significance = 6.33
 Residue Identity = 38% Matches = 14 Mismatches = 13
 Gaps = 9 Conservative Substitutions = 0

GLFAYKAATKAGYCGGAVLAKDGADTFIVGTHSAGGNGVGYCSCVSRSMLLKMKAHIDPEPHHEGLIVDTRD
 1800 1810 1820 1830 1840 1850 1860 1870

X 10 20 X
 KAL-----PVVLENARILKNCVDAKMTEEDKE
 || || | |||||
 VEERVHVMRKTKLAPTVAHGIVENPEFGPAALSNKDPRLNEGVLDEVIFSKHKGDTKMTEEDKALFRRCAAD
 1880 1890 1900 1910 1920 1930 X 1940

YASRLHNVLGTANAPLSIYEAIKQVDGLDAMEPDTAPGLPWALQGKRRGTLIDFENGTVGPEVASALELMEK
 1950 1960 1970 1980 1990 2000 2010

RQYKFTCQTFLKDEVPRMEK
 2020 2030

6. US-08-300-510-2 (1-27)
 S40064 3-deoxy-D-manno-2-octulosonic acid (Kdo) transferase

ENTRY S40064 #type complete
 TITLE 3-deoxy-D-manno-2-octulosonic acid (Kdo) transferase -
 Chlamydia psittaci
 ORGANISM #formal_name Chlamydia psittaci
 DATE 19-May-1994; #sequence_revision 19-May-1994; #text_change
 19-May-1994
 ACCESSIONS S40064
 REFERENCE S40064
 #authors Mamat, U.; Baumann, M.; Schmidt, G.; Brade, H.
 #journal Mol. Microbiol. (1993) 10:935-941
 #title The genus-specific lipopolysaccharide epitope of Chlamydia is
 assembled in C. psittaci and C. trachomatis by

glycosyltransferases of low homology.

#accession S40064
##status preliminary
##residues 1-411 ##label MAM
##cross-references EMBL:X69476
SUMMARY #length 411 #molecular-weight 46618 #checksum 5839
SEQUENCE

Initial Score = 11 Optimized Score = 11 Significance = 5.62
Residue Identity = 40% Matches = 11 Mismatches = 16
Gaps = 0 Conservative Substitutions = 0

AVIINGKLSANSCKRFTILKRFRNYFSPVDGFLQDEQHKARFLQLGVDKEKIQVTGNIKTYTETLSENNQ
140 150 160 170 180 190 200
X 10 20 X
KALPVVLENARILKNCVDAKMTEEDKE
||||| | || | ||
RDYWREKLQLAQDTELLVLGSHPKDVEVWLPVVRELRRNLKVLWVPRHIERSKELEALLSKENISYGLWSK
210 220 230 240 250 260 X 270 280
EATFAQHDAAIIVDAIGWLKQLYSAADLAFVGGTFDDRIGGHNLEPLQCGVPLIFGPHIQSQSDLAERLLSM
290 300 310 320 330 340 350
GAGCCLDKTNI
360

7. US-08-300-510-2 (1-27)

S28562 3-deoxy-D-manno-2-octulosonic acid (Kdo) transferase

ENTRY S28562 #type complete
TITLE 3-deoxy-D-manno-2-octulosonic acid (Kdo) transferase -
Chlamydia psittaci
ORGANISM #formal_name Chlamydia psittaci
DATE 12-Mar-1993 #sequence_revision 12-Mar-1993 #text_change
18-Jun-1993
ACCESSIONS S28562
REFERENCE S28562
#authors Mamat, U.; Baumann, M.; Schmidt, G.; Brade, H.
#submission submitted to the EMBL Data Library, November 1992
#description Cloning and sequence analysis of a Chlamydia psittaci gene
involved in the expression of the genus-specific
lipopolysaccharide epitope.
#accession S28562
##molecule_type DNA
##residues 1-411 ##label MAM
##cross-references EMBL:X69476

GENETICS

#gene gseA
SUMMARY #length 411 #molecular-weight 46618 #checksum 5839
SEQUENCE

Initial Score = 11 Optimized Score = 11 Significance = 5.62
Residue Identity = 40% Matches = 11 Mismatches = 16
Gaps = 0 Conservative Substitutions = 0

AVIINGKLSANSCKRFTILKRFRNYFSPVDGFLQDEQHKARFLQLGVDKEKIQVTGNIKTYTETLSENNQ
140 150 160 170 180 190 200
X 10 20 X
KALPVVLENARILKNCVDAKMTEEDKE
||||| | || | ||
RDYWREKLQLAQDTELLVLGSHPKDVEVWLPVVRELRRNLKVLWVPRHIERSKELEALLSKENISYGLWSK
210 220 230 240 250 260 X 270 280

EATFAWDAIIVDAIGWLRGLYSADLAFVGGTFDDRIIGGHNLEPLQCGVPLIFGPHIWSWSDLERLLSH
 290 300 310 320 330 340 350

GAGCCLDKNTI
 360

8. US-08-300-510-2 (1-27)

JN0431 RNA-directed RNA polymerase (EC 2.7.7.48) - foot-a

ENTRY JN0431 #type complete
 TITLE RNA-directed RNA polymerase (EC 2.7.7.48) - foot-and-mouth
 disease virus A (strain A22)
 ORGANISM #formal_name Aphthovirus A #common_name foot-and-mouth
 disease virus A
 DATE 05-Mar-1993 #sequence_revision 05-Mar-1993 #text_change
 30-Sep-1993
 ACCESSIONS JN0431
 REFERENCE JN0431
 #authors Kuzmin, I.V.; Rybakov, S.S.; Ivanyushchenkov, V.N.; Burdov,
 A.N.
 #journal Bioorg. Khim. (1989) 15:419-422
 #title Nucleotide sequence of the FMDV A22 RNA polymerase gene.
 #cross-references MUID:89302183
 #accession JN0431
 ##molecule_type mRNA
 ##residues 1-470 ##label KUZ
 ##note this paper is in Russian, with an English abstract
 CLASSIFICATION #superfamily foot-and-mouth disease virus genome polyprotein
 KEYWORDS nucleotidyltransferase
 SUMMARY #length 470 #molecular-weight 52657 #checksum 1182
 SEQUENCE

Initial Score = 11 Optimized Score = 12 Significance = 5.62
 Residue Identity = 36% Matches = 13 Mismatches = 14
 Gaps = 9 Conservative Substitutions = 0

```

                                X          10          20          X
                                KAL-----PVVLENARILKNCVDAKMTEEDKE
                                ||          ||          |          |          ||
GLIVDTRDVEERVHVMRKTKLAPTVAHG VFNPEFGPAALFNKDPRLNEGVVLDEVIFSKHKGDTKMTAEDKA
      10          20          30          X 40          50          60          70 X

LFRACAADYASRLHNVLTANAPLSIYEAIKGIDGLDAMEPDTAPGLPWALQGQRRGALIDFENGTVGPEVA
      80          90          100          110          120          130          140

SALELMEKRQYKFTCQTLKDEV RPMEK
      150          160          170

```

9. US-08-300-510-2 (1-27)

GNNYF genome polyprotein - foot-and-mouth disease virus

ENTRY GNNYF #type complete
 TITLE genome polyprotein - foot-and-mouth disease virus O (strains
 01K and 01BFS)
 CONTAINS coat protein VP1; coat protein VP2; coat protein VP3; coat
 protein VP4; core protein p12; core protein p14; core
 protein P20b; core protein p34; core protein P56; core
 protein VPg; nonstructural protein p20a
 ORGANISM #formal_name Aphthovirus O #common_name foot-and-mouth
 disease virus O
 #note host Artiodactyla (cloven-footed mammals)
 DATE 01-Sep-1981 #sequence_revision 27-Nov-1985 #text_change
 08-Apr-1994
 ACCESSIONS A03907; A37503

REFERENCE A03907
 #authors Forss, S.; Strebel, K.; Beck, E.; Schaller, H.
 #journal Nucleic Acids Res. (1984) 12:6587-6601
 #title Nucleotide sequence and genome organization of foot-and-mouth disease virus.
 #cross-references MUID:84297249
 #contents strain 01K
 #accession A03907
 ##molecule_type mRNA
 ##residues 1-2332 ##label FOR

REFERENCE A37503
 #authors Makoff, A.J.; Paynter, C.A.; Rowlands, D.J.; Boothroyd, J.C.
 #journal Nucleic Acids Res. (1982) 10:8285-8295
 #title Comparison of the amino acid sequence of the major immunogen from three serotypes of foot and mouth disease virus.
 #cross-references MUID:83143292
 #contents strain 01BFS
 #accession A37503

##molecule_type genomic RNA
 ##residues 715-779,'V',781-807,'R',809-860,'S',862-951 ##label MAK

COMMENT The coat protein VP1 contains the main antigenic determinants of the virion; therefore, changes in its sequence must be responsible for the high antigenic variability of the virus.

COMMENT Coat proteins VP2 and VP3 are related to the poliovirus coat proteins VP2 and VP3.

CLASSIFICATION #superfamily foot-and-mouth disease virus genome polyprotein
 KEYWORDS coat protein; core protein; nonstructural protein; polyprotein

FEATURE

1-217 #product nonstructural protein p20a #label NPA\
 218-286 #product coat protein VP4 #label VP4\
 287-504 #product coat protein VP2 #label VP2\
 505-724 #product coat protein VP3 #label VP3\
 725-937 #product coat protein VP1 #label VP1\
 938-1107 #product core protein p12 #label C12\
 1108-1425 #product core protein p34 #label P34\
 1426-1578 #product core protein p14 #label C14\
 1579-1649 #product genome-linked protein VPg #label VPG\
 1650-1862 #product nonstructural protein p20b #label P20\
 1863-2332 #product RNA-directed RNA polymerase #label P56

SUMMARY #length 2332 #molecular-weight 258925 #checksum 4170
 SEQUENCE

Initial Score = 11 Optimized Score = 12 Significance = 5.62
 Residue Identity = 36% Matches = 13 Mismatches = 14
 Gaps = 9 Conservative Substitutions = 0

GLFAYRAATKAGYCGGAVLAKDGADTFIVGTHSAGGNGVGYCSCVSRSMLLKMKAHIDPEPHHEGLIVDTRD
 1800 1810 1820 1830 1840 1850 1860 1870

X 10 20 X
 KAL-----PVVLENARILKNCVDAKMTEEDKE
 || ||| | ||||
 VEERVHVMRKTKLAPTVAHG VFNPFCGAALSNKDPRLNEG VVLDEVIFSKHKGDTKMSEEDKALFRRCAAD
 1880 1890 1900 1910 1920 1930 X 1940

YASRLHSVLGTANAPLSIYEAIKGVDGLDAMEPDTAPGLPWALQKRRGALIDFENGTVGPEVEAALKLMEK
 1950 1960 1970 1980 1990 2000 2010

REYKFVCQTFLEKDEIRPLEK
 2020 2030

ENTRY S02068 #type complete
 TITLE RNA-directed RNA polymerase (EC 2.7.7.48) - foot-and-mouth disease virus A
 ALTERNATE_NAMES RNA replicase
 ORGANISM #formal_name Aphthovirus A #common_name foot-and-mouth disease virus A
 DATE 01-Dec-1989 #sequence_revision 01-Dec-1989 #text_change 30-Sep-1993
 ACCESSIONS S02068
 REFERENCE S02068
 #authors Villaverde, A.; Martinez-Salas, E.; Domingo, E.
 #journal J. Mol. Biol. (1988) 204:771-776
 #title 3D gene of foot-and-mouth disease virus. Conservation by convergence of average sequences.
 #cross-references MUID:89141768
 #accession S02068
 ##molecule_type mRNA
 ##residues 1-470 ##label VIL
 ##note 48-Gly, 68-Ala, 158-Val, 274-Ile, 306-Ile, 374-Leu, and 444-Glu were also found
 ##note sequence not compared to nucleotide translation
 GENETICS
 #gene 3D
 CLASSIFICATION #superfamily foot-and-mouth disease virus genome polyprotein
 KEYWORDS nucleotidyltransferase
 SUMMARY #length 470 #molecular-weight 52910 #checksum 502
 SEQUENCE

Initial Score = 10 Optimized Score = 11 Significance = 4.92
 Residue Identity = 33% Matches = 12 Mismatches = 15
 Gaps = 9 Conservative Substitutions = 0

```

                                X          10          20          X
                                KAL-----PVVLENARILKNCVDAKMTEEDKE
                                ||          |||          | || ||||
GLIVDTRDVEERVHVMRKTKLAPTVAHG VFNP EFGPAALSNKDPRLNEG VVLDEVIFSRHKGDTKMSEEDKA
      10          20          30          X 40          50          60          70 X

LFRRCAADYASRLHSVLGTANAPLSIYEAIKGV DGLDAMEPDTAPGLPWALQ GKRRGALIDFENGTVGPEAE
      80          90          100          110          120          130          140

AALKLMEKREYKFACQTF LKDEIRPMEK
      150          160          170

```

11. US-08-300-510-2 (1-27)

S10340 DNA-directed RNA polymerase (EC 2.7.7.6) - yeast

ENTRY S10340 #type complete
 TITLE DNA-directed RNA polymerase (EC 2.7.7.6) - yeast (Kluyveromyces marxianus var. lactis)
 ORGANISM #formal_name Kluyveromyces marxianus var. lactis, Candida sphaerica
 DATE 21-Nov-1993; #sequence_revision 21-Nov-1993; #text_change 21-Nov-1993
 ACCESSIONS S10340
 REFERENCE S10336
 #authors Wilson, D.W.; Meacock, P.A.
 #journal Nucleic Acids Res. (1988) 16:8097-8112
 #title Extranuclear gene expression in yeast: evidence for a plasmid-encoded RNA polymerase of unique structure.
 #cross-references MUID:88335549
 #accession S10340
 ##status preliminary
 ##residues 1-982 ##label WIL
 ##cross-references EMBL:X07946

SUMMARY #length 982 #molecular-weight 113960 #checksum 5412
SEQUENCE

Initial Score = 10 Optimized Score = 10 Significance = 4.92
Residue Identity = 37% Matches = 10 Mismatches = 17
Gaps = 0 Conservative Substitutions = 0

DILIGLGAWNTIKEIWSIDRSKIKIDSKTGRINWIRYDKEMEIGQYFKICLSYMRS�GRDILIKNDKYSIVE
640 650 660 670 680 690 700

X 10 20 X
KALPVVLENARILKNCVDAKMTEEDKE

FDNSYLPKTDTMKFGDLVDLRIYKGIVMLPLCLRSTYLNKLYVDRKYSEAEKEVTKLLKSKNGAYHTLVE
710 720 730 X 740 750 760 X 770

GHRVDR CIRSVIVPDTL DIDTIKIPFGANIGCEYGLLN RQPSLN VDSIKLVK LKQGSNKTIAINPLLCQSF
780 790 800 810 820 830 840 850

NADFDGDEMNI
860

12. US-08-300-510-2 (1-27)

S00964 hypothetical protein 6 - yeast (Kluyveromyces marx

ENTRY S00964 #type complete
TITLE hypothetical protein 6 - yeast (Kluyveromyces marxianus var.
lactis) plasmid pGK12
ORGANISM #formal_name Kluyveromyces marxianus var. lactis, Candida
sphaerica
DATE 30-Sep-1989 #sequence_revision 30-Sep-1989 #text_change
18-Jun-1993
ACCESSIONS S00964
REFERENCE S00959
#authors Tommasino, M.; Ricci, S.; Galeotti, C.L.
#journal Nucleic Acids Res. (1988) 16:5863-5878
#title Genome organization of the killer plasmid pGK12 from
Kluyveromyces lactis.
#cross-references MUID:88289339
#accession S00964
##molecule_type DNA
##residues 1-982 ##label TOM
##cross-references EMBL:X07776

GENETICS

#genome plasmid

SUMMARY #length 982 #molecular-weight 113960 #checksum 5276
SEQUENCE

Initial Score = 10 Optimized Score = 10 Significance = 4.92
Residue Identity = 37% Matches = 10 Mismatches = 17
Gaps = 0 Conservative Substitutions = 0

DILIGLGAWNTIKEIWSIDRSKIKIDSKTGRINWIRYDKEMEIGQYFKICLSYMRS�GRDILIKNDKYSIVE
640 650 660 670 680 690 700

X 10 20 X
KALPVVLENARILKNCVDAKMTEEDKE

FDNSYLPKTDTMKFGDLVDLRIYKGIVMLPLCLRSTYLNKLYVDRKYSEAEKEVTKLLKSKNGAYHTLVE
710 720 730 X 740 750 760 X 770

GHRVDR CIRSVIVPDTL DIDTIKIPFGANIGCEYGLLN RQPSLN VDSIKLVK LKQGSNKTIAINPLLCQSF
780 790 800 810 820 830 840 850

NADFDGDEMNI

13. US-08-300-510-2 (1-27)

GNNY4F genome polyprotein - foot-and-mouth disease virus

ENTRY GNNY4F #type complete
 TITLE genome polyprotein - foot-and-mouth disease virus A (strain A12)
 CONTAINS coat protein VP1; coat protein VP2; coat protein VP3; coat protein VP4; core protein p14; core protein p19; core protein p41; core protein X; genome-linked protein VPg1; genome-linked protein VPg2; genome-linked protein VPg3; nonstructural protein p20a; proteinase (EC 3.4.-.-); RNA-directed RNA polymerase (EC 2.7.7.48)
 ORGANISM #formal_name Aphthovirus A #common_name foot-and-mouth disease virus A
 DATE 30-Sep-1987 #sequence_revision 30-Sep-1987 #text_change 08-Apr-1994
 ACCESSIONS A25794
 REFERENCE A25794
 #authors Robertson, B.H.; Grubman, M.J.; Weddell, G.N.; Moore, D.M.; Welsh, J.D.; Fischer, T.; Dowbenko, D.J.; Yansura, D.G.; Small, B.; Kleid, D.G.
 #journal J. Virol. (1985) 54:651-660
 #title Nucleotide and amino acid sequence coding for polypeptides of foot-and-mouth disease virus type A12.
 #cross-references MUID:85211015
 #accession A25794
 ##molecule_type genomic RNA
 ##residues 1-2332 ##label ROB
 CLASSIFICATION #superfamily foot-and-mouth disease virus genome polyprotein
 KEYWORDS coat protein; core protein; genome-linked protein; hydrolase; nonstructural protein; nucleotidyltransferase; polyprotein; proteinase

FEATURE
 1-216 #product nonstructural protein p20a #label NPA\
 217-285 #product coat protein VP4 #label VP4\
 286-503 #product coat protein VP2 #label VP2\
 504-723 #product coat protein VP3 #label VP3\
 724-937 #product coat protein VP1 #label VP1\
 938-953 #product core protein X #label CPX\
 954-1107 #product core protein p14 #label C14\
 1108-1425 #product core protein p41 #label C41\
 1426-1578 #product core protein p19 #label C19\
 1579-1601 #product genome-linked protein VPg1 #label VG1\
 1602-1625 #product genome-linked protein VPg2 #label VG2\
 1626-1649 #product genome-linked protein VPg3 #label VG3\
 1650-1862 #product proteinase #label PTS\
 1863-2332 #product RNA-directed RNA polymerase #label RRP
 SUMMARY #length 2332 #molecular-weight 259408 #checksum 6669
 SEQUENCE

Initial Score = 10 Optimized Score = 11 Significance = 4.92
 Residue Identity = 33% Matches = 12 Mismatches = 15
 Gaps = 9 Conservative Substitutions = 0

SLFAYKAATKAGYCGGAVLAKDGADTFIVGTHSAGCGNGVGYCSCVSKSMLLRMKAHVDPEPQHEGLIVDTRD
 1800 1810 1820 1830 1840 1850 1860 1870
 X 10 20 X
 KAL-----PVVLENARILKNCVDAKMTEEDKE
 || ||| | ||| |||
 VEERVHVHRKTKLAPTVAHG VFNPEFGPAALSNKDPRLNEGVVLDEVIFSKHKGDTKMSAEDKALFRACAAAD
 1880 1890 1900 1910 1920 1930 X 1940

YASRLHSVLGTANAFLESTIETATRGVDGLDAMESDTAFGLFWAFGRKRRGALIDFENGIVGFVEVAILLHMER
 1950 1960 1970 1980 1990 2000 2010

REYKFVCQTFLKDEIRPMEK
 2020 2030

14. US-08-300-510-2 (1-27)
 C30305 submandibular gland protein (spot 1) precursor - r

ENTRY C30305 #type complete
 TITLE submandibular gland protein (spot 1) precursor - rat
 ORGANISM #formal_name Rattus norvegicus #common_name Norway rat
 DATE 22-Nov-1989 #sequence_revision 22-Nov-1989 #text_change
 17-Feb-1994
 ACCESSIONS C30305
 REFERENCE A30305
 #authors Dickinson, D.P.; Mirels, L.; Tabak, L.A.; Gross, K.W.
 #journal Mol. Biol. Evol. (1989) 6:80-102
 #title Rapid evolution of variants in a rodent multigene family
 encoding salivary proteins.
 #cross-references MUID:89158788
 #accession C30305
 ##status preliminary
 ##molecule_type mRNA
 ##residues 1-91 ##label DIC
 CLASSIFICATION #superfamily submandibular gland 16.5K protein
 SUMMARY #length 91 #molecular-weight 9227 #checksum 7778
 SEQUENCE

Initial Score = 9 Optimized Score = 10 Significance = 4.22
 Residue Identity = 37% Matches = 10 Mismatches = 17
 Gaps = 0 Conservative Substitutions = 0

X 10 20 X
 KALPVVLENARILKNCVDAKMTEEDKE
 || || || || || || ||
 MKFLALLVLLGVSTILVSCQDAGTDTADTSDTADGTTDSGTQADATDGGQDAESSDGTSDAVDGDAPAEQDQ
 X 10 20 30 40 50 60 70
 EDSALLALVNTLKEKFTLG
 80 90

15. US-08-300-510-2 (1-27)
 A35072 nonhistone chromosomal protein NHP6A - yeast (Sacc

ENTRY A35072 #type complete
 TITLE nonhistone chromosomal protein NHP6A - yeast (Saccharomyces
 cerevisiae)
 ORGANISM #formal_name Saccharomyces cerevisiae
 DATE 22-Jan-1993 #sequence_revision 22-Jan-1993 #text_change
 06-May-1994
 ACCESSIONS A35072; S31260; C44031
 REFERENCE A35072
 #authors Kolodrubetz, D.; Burgum, A.
 #journal J. Biol. Chem. (1990) 265:3234-3239
 #title Duplicated NHP6 genes of Saccharomyces cerevisiae encode
 proteins homologous to bovine high mobility group protein
 1.
 #cross-references MUID:90153974
 #accession A35072
 ##molecule_type DNA
 ##residues 1-93 ##label KOL
 ##cross-references EMBL:X15317
 REFERENCE A44031

#authors Terceiro, J.C.; Riles, L.E.; Wilkner, R.D.
 #journal J. Biol. Chem. (1992) 267:20270-20276
 #title Localized mutagenesis and evidence for post-transcriptional regulation of MAK3. A putative N-acetyltransferase required for double-stranded RNA virus propagation in Saccharomyces cerevisiae.
 #cross-references MUID:93015901
 #accession S31260
 ##molecule_type DNA
 ##residues 1-93 ##label TER
 ##cross-references EMBL:M95912

GENETICS

#gene LISTA:NHP6A
 #map_position 16
 CLASSIFICATION #superfamily nonhistone chromosomal protein HMG-2; HMG box homology
 KEYWORDS chromosomal protein

FEATURE

18-93 #domain HMG box homology #label HMG1
 SUMMARY #length 93 #molecular-weight 10802 #checksum 4901
 SEQUENCE

Initial Score	=	9	Optimized Score	=	9	Significance	=	4.22
Residue Identity	=	33%	Matches	=	9	Mismatches	=	18
Gaps	=	0	Conservative Substitutions	=			=	0

X 10
KALPVVLENARIL
|||

MVTPREPKKRTTRKKKDPNAPKRALSAYMFFANENRDIVRSENPDITFGQVGKKLGEKWKALTPEEKQPYEA
 10 20 30 40 50 60 70

20 X
KNCVDAKMTEEDKE
| | | | |
KAQADKKRYESEKELYNATLA
80 X 90

> 0 <
 0| 0 IntelliGenetics
 > 0 <

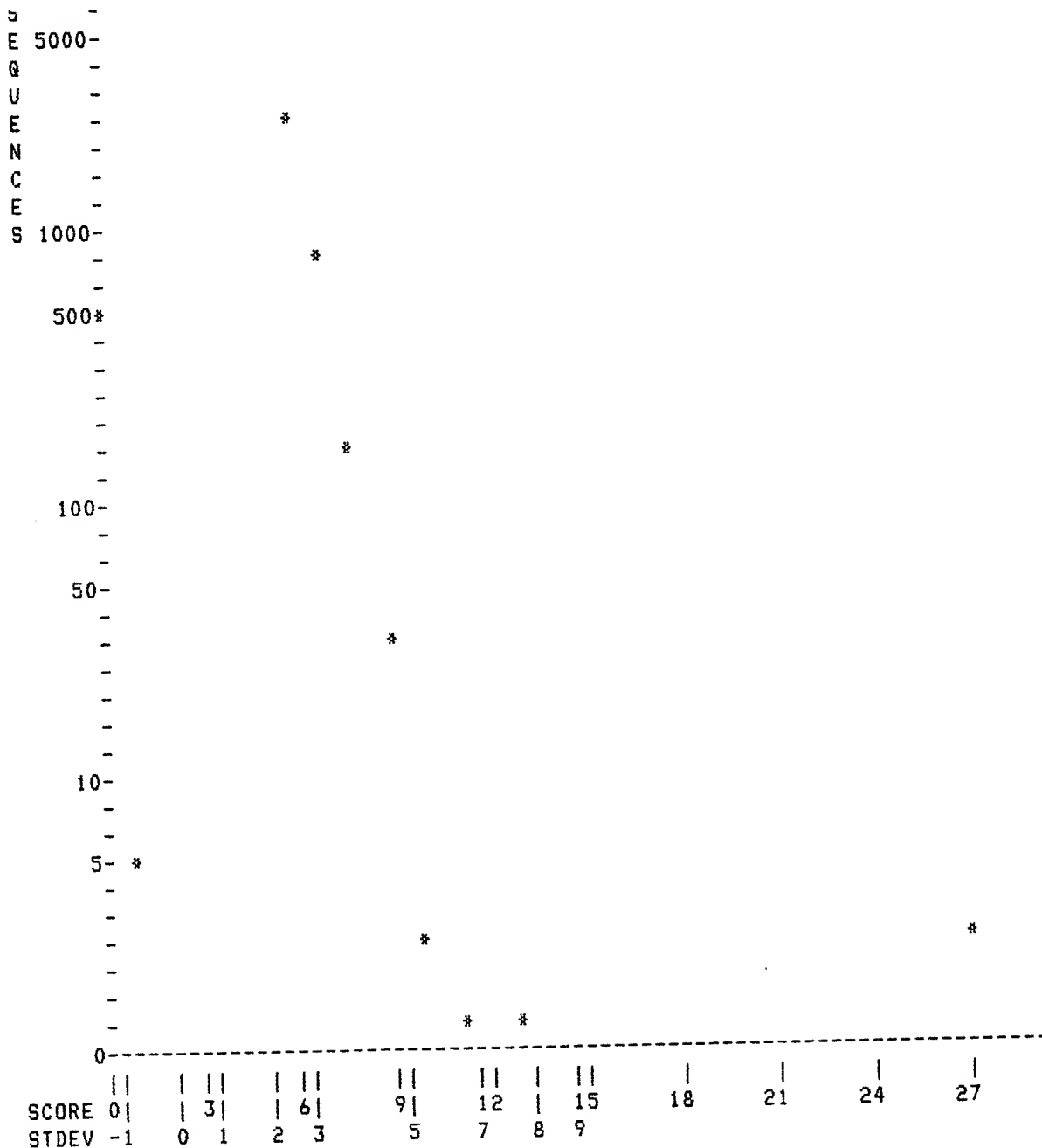
FastDB - Fast Pairwise Comparison of Sequences
 Release 5.4

Results file 2-spt.res made by on Fri 24 Mar 95 7:54:42-PST.

Query sequence being compared:US-08-300-510-2 (1-27)
 Number of sequences searched: 40292
 Number of scores above cutoff: 3966

Results of the initial comparison of US-08-300-510-2 (1-27) with:
 Data bank : Swiss-Prot 30, all entries

100000-
 -
 N -
 U50000-
 M -
 B -
 E -
 R -
 -
 D - *
 F10000- * *
 - *



PARAMETERS

Similarity matrix	Unitary	K-tuple	2
Mismatch penalty	1	Joining penalty	20
Gap penalty	1.00	Window size	27
Gap size penalty	0.05		
Cutoff score	0		
Randomization group	0		
Initial scores to save	40	Alignments to save	15
Optimized scores to save	0	Display context	100

SEARCH STATISTICS

Scores:	Mean	Median	Standard Deviation
	3	5	1.34
Times:	CPU	Total Elapsed	
	00:00:36.14	00:00:38.00	

Number of residues:	14147368
Number of sequences searched:	40292

Cut-off raised to 3.
 Cut-off raised to 4.
 Cut-off raised to 5.
 Cut-off raised to 6.
 Cut-off raised to 7.

The scores below are sorted by initial score.
 Significance is calculated based on initial score.

2 100% similar sequences to the query sequence were found:

Sequence Name	Description	Length	Init. Opt.		Sig.	Frame
			Score	Score		
1. FELA_FELCA	MAJOR ALLERGEN I POLYPEPTIDE	92	27	27	17.85	0
2. FELB_FELCA	MAJOR ALLERGEN I POLYPEPTIDE	88	27	27	17.85	0

The list of other best scores is:

Sequence Name	Description	Length	Init.	Opt.	Score	Sig.	Frame

**** 7 standard deviations above mean ****							
3. POLG_FMDV1	GENOME POLYPROTEIN (NONSTRUCT	2333	13	14	7.44	0	
**** 5 standard deviations above mean ****							
4. POLG_FMDV0	GENOME POLYPROTEIN (NONSTRUCT	2332	11	12	5.95	0	
5. RPOL_KLULA	PROBABLE DNA-DIRECTED RNA POL	982	10	10	5.21	0	
6. POLG_FMDVA	GENOME POLYPROTEIN (NONSTRUCT	2332	10	11	5.21	0	
**** 4 standard deviations above mean ****							
7. NHPA_YEAST	NONHISTONE CHROMOSOMAL PROTEI	93	9	9	4.46	0	
8. NHPB_YEAST	NONHISTONE CHROMOSOMAL PROTEI	99	9	9	4.46	0	
9. HEMM_THEZO	MYOHENERYTHRIN.	118	9	11	4.46	0	
10. PHEA_FREDI	C-PHYCOERYTHRIN ALPHA CHAIN.	164	9	10	4.46	0	
11. ATPF_PROMO	ATP SYNTHASE B CHAIN, SODIUM	168	9	11	4.46	0	
12. UCRI_SYNP2	CYTOCHROME B6-F COMPLEX IRON-	180	9	9	4.46	0	
13. XYNB_STRLI	ENDO-1,4-BETA-XYLANASE B PREC	333	9	9	4.46	0	
14. YB09_YEAST	HYPOTHETICAL 38.7 KD PROTEIN	347	9	9	4.46	0	
15. ADH4_YEAST	ALCOHOL DEHYDROGENASE IV (EC	382	9	9	4.46	0	
16. PGKB_TRYBB	PHOSPHOGLYCERATE KINASE, CYTO	421	9	11	4.46	0	
17. PRO1_LISMO	ZINC METALLOPROTEINASE PRECUR	510	9	10	4.46	0	
18. VNUC_INCCA	NUCLEOPROTEIN.	565	9	9	4.46	0	
19. PRIM_BPT7	DNA PRIMASE, CHAINS A AND B (566	9	10	4.46	0	
20. PRIM_BPT3	DNA PRIMASE (EC 2.7.7.-).	566	9	10	4.46	0	
21. CEIB_ECOLI	COLICIN IB PROTEIN.	626	9	9	4.46	0	
22. CEIA_ECOLI	COLICIN IA PROTEIN.	626	9	9	4.46	0	
23. CRAC_DICDI	PROTEIN CRAC.	698	9	10	4.46	0	
24. LON_BACBR	ATP-DEPENDENT PROTEASE LA (EC	779	9	9	4.46	0	
25. YKS8_YEAST	70 KD PEROXISOMAL MEMBRANE PR	853	9	10	4.46	0	
26. POLG_FMDVS	GENOME POLYPROTEIN (COAT PROT	861	9	10	4.46	0	
27. FOX2_YEAST	PEROXISOMAL HYDRATASE-DEHYDRO	900	9	10	4.46	0	
28. BVGC_BORPE	SENSOR PROTEIN BVGC (EC 2.7.3	936	9	9	4.46	0	
29. SYI_YEAST	ISOLEUCYL-TRNA SYNTHETASE (EC	1072	9	9	4.46	0	
30. BVGS_BORBR	VIRULENCE SENSOR PROTEIN BVGS	1238	9	9	4.46	0	
31. NOS1_RAT	NITRIC-OXIDE SYNTHASE, BRAIN	1429	9	9	4.46	0	
32. SSP5_STRSA	AGGLUTININ RECEPTOR PRECURSOR	1473	9	10	4.46	0	
**** 3 standard deviations above mean ****							
33. NIFW_KLEPN	NIFW PROTEIN.	86	8	8	3.72	0	
34. RL12_EUGGR	50S RIBOSOMAL PROTEIN L12.	131	8	9	3.72	0	
35. YBEG_ECOLI	HYPOTHETICAL 18.4 KD PROTEIN	156	8	8	3.72	0	
36. RL12_SPIOL	50S RIBOSOMAL PROTEIN L12, CH	189	8	9	3.72	0	
37. INA5_HUMAN	INTERFERON ALPHA-5 PRECURSOR	189	8	9	3.72	0	
38. A1AG_RABIT	ALPHA-1-ACID GLYCOPROTEIN PRE	201	8	9	3.72	0	
39. RIP3_SAPOF	RIBOSOME-INACTIVATING PROTEIN	236	8	11	3.72	0	

1. US-08-300-510-2 (1-27)

FELA_FELCA MAJOR ALLERGEN I POLYPEPTIDE CHAIN 1 MAJOR FORM PR

ID FELA_FELCA STANDARD; PRT; 92 AA.
AC P30438;
DT 01-APR-1993 (REL. 25, CREATED)
DT 01-APR-1993 (REL. 25, LAST SEQUENCE UPDATE)
DT 01-JUN-1994 (REL. 29, LAST ANNOTATION UPDATE)
DE MAJOR ALLERGEN I POLYPEPTIDE CHAIN 1 MAJOR FORM PRECURSOR (FEL D I)
DE (CAT-1) (AG 4).
GN CH1.
OS FELIS CATUS (CAT).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; CARNIVORA.
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 23-92.
RC TISSUE=SALIVARY GLAND;
RM 92052157
RA MORGENSTERN J.P., GRIFFITH I.J., BRAUER A.W., ROGERS B.L.,
RA BOND J.F., CHAPMAN M.D., KUO M.-C.;
RL PROC. NATL. ACAD. SCI. U.S.A. 88:9690-9694(1991).
RN [2]
RP SEQUENCE FROM N.A.
RM 92241678
RA GRIFFITH I.J., CRAIG S., POLLOCK J., YU X.-B., MORGENSTERN J.P.,
RA ROGERS B.L.,
RL GENE 113:263-268(1992).
RN [3]
RP SEQUENCE OF 23-62, AND CHARACTERIZATION.
RM 91287714
RA DUFFORT D.A., CARREIRA J., NITTI G., POLO F., LOMBARDERO M.;
RL MOL. IMMUNOL. 28:301-309(1991).
RN [4]
RP CHARACTERIZATION.
RA LEITERMANN K., OHMAN J.L. JR.;
RL J. ALLERGY CLIN. IMMUNOL. 74:147-153(1991).
CC -!- DISEASE: MAJOR ALLERGEN PRODUCED BY THE DOMESTIC CAT.
CC -!- SUBUNIT: HETEROTETRAMER COMPOSED OF TWO NON-COVALENTLY LINKED
CC DISULFIDE-LINKED HETERODIMER OF CHAINS 1 AND 2.
CC -!- TISSUE SPECIFICITY: SALIVA, AND SEBACEOUS GLANDS.
CC -!- ALTERNATIVE PRODUCTS: USAGE OF TWO DIFFERENT INITIATOR MET ARE
CC RESPONSIBLE FOR THE PRODUCTION OF TWO FORMS OF THE SIGNAL SEQUENCE
CC OF THIS ALLERGEN SUBUNIT.
CC -!- SIMILARITY: TO UTEROGLOBIN.
DR EMBL; M74952; FDFELD1.
DR PIR; JC1136; JC1136.
DR PROSITE; PS00403; UTEROGLOBIN_1.
DR PROSITE; PS00404; UTEROGLOBIN_2.
KW ALLERGEN; SIGNAL; ALTERNATIVE SPLICING.
FT SIGNAL 1 22
FT CHAIN 23 92 MAJOR ALLERGEN I POLYPEPTIDE CHAIN 1.
FT DISULFID 25 25 INTERCHAIN (POTENTIAL).
FT DISULFID 92 92 INTERCHAIN (POTENTIAL).
FT VARIANT 51 51 K -> N.
FT CONFLICT 5 5 R -> C (IN REF. 2).
FT CONFLICT 18 18 W -> S (IN REF. 2).
FT CONFLICT 82 82 L -> V (IN REF. 2).
SQ SEQUENCE 92 AA; 10252 MW; 43206 CN;

Initial Score = 27 Optimized Score = 27 Significance = 17.85
Residue Identity = 100% Matches = 27 Mismatches = 0
Gaps = 0 Conservative Substitutions = 0

KALPVVLENARILKNCVDAKMT
 |||||

MKGARVLVLLWAALLLIWGGNCEICPAVKRBDVDFLTGTPDEYVEQVAQYKALPVVLENARILKNCVDAKMT

10 20 30 40 50 60 70

X
 EEDKE
 |||||
 EEDKENALSLLDKIYTSPLC
 X 80 90

2. US-08-300-510-2 (1-27)

FELB_FELCA MAJOR ALLERGEN I POLYPEPTIDE CHAIN 1 MINOR FORM PR

ID FELB_FELCA STANDARD; PRT: 88 AA.
 AC P30439;
 DT 01-APR-1993 (REL. 25, CREATED)
 DT 01-APR-1993 (REL. 25, LAST SEQUENCE UPDATE)
 DT 01-JUN-1994 (REL. 29, LAST ANNOTATION UPDATE)
 DE MAJOR ALLERGEN I POLYPEPTIDE CHAIN 1 MINOR FORM PRECURSOR (FEL D I)
 DE (CAT-1) (AG 4).
 GN CH1.
 OS FELIS CATUS (CAT).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
 OC EUTHERIA; CARNIVORA.
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 19-88.
 RM 92052157
 RA MORGENSTERN J.P., GRIFFITH I.J., BRAUER A.W., ROGERS B.L.,
 RA BOND J.F., CHAPMAN M.D., KUO M.-C.;
 RL PROC. NATL. ACAD. SCI. U.S.A. 88:9690-9694(1991).
 RN [2]
 RP SEQUENCE FROM N.A.
 RM 92241678
 RA GRIFFITH I.J., CRAIG S., POLLOCK J., YU X.-B., MORGENSTERN J.P.,
 RA ROGERS B.L.,
 RL GENE 113:263-268(1992).
 RN [3]
 RP SEQUENCE OF 19-58, AND CHARACTERIZATION.
 RM 91287714
 RA DUFFORT O.A., CARREIRA J., NITTI G., POLO F., LOMBARDERO M.;
 RL MOL. IMMUNOL. 28:301-309(1991).
 RN [4]
 RP CHARACTERIZATION.
 RA LEITERMANN K., OHMAN J.L. JR.;
 RL J. ALLERGY CLIN. IMMUNOL. 74:147-153(1991).
 CC -!- DISEASE: MAJOR ALLERGEN PRODUCED BY THE DOMESTIC CAT.
 CC -!- SUBUNIT: HETEROTETRAMER COMPOSED OF TWO NON-COVALENTLY LINKED
 CC DISULFIDE-LINKED HETERODIMER OF CHAINS 1 AND 2.
 CC -!- TISSUE SPECIFICITY: SALIVA, AND SEBACEOUS GLANDS.
 CC -!- ALTERNATIVE PRODUCTS: USAGE OF TWO DIFFERENT INITIATOR MET ARE
 CC RESPONSIBLE FOR THE PRODUCTION OF TWO FORMS OF THE SIGNAL SEQUENCE
 CC OF THIS ALLERGEN SUBUNIT.
 CC -!- SIMILARITY: TO UTEROGLOBIN.
 DR EMBL: M74953; FDFELDIB.
 DR PIR: JC1126; JC1126.
 DR PROSITE: PS00403; UTEROGLOBIN_1.
 DR PROSITE: PS00404; UTEROGLOBIN_2.
 KW ALLERGEN; SIGNAL; ALTERNATIVE SPLICING.

FT	SIGNAL	1	18	
FT	CHAIN	19	88	MAJOR ALLERGEN I POLYPEPTIDE CHAIN 1.
FT	DISULFID	21	21	INTERCHAIN (POTENTIAL).
FT	DISULFID	88	88	INTERCHAIN (POTENTIAL).
FT	VARIANT	47	47	K -> N.

FI CONFLICT 78 78 L -> V (IN REF. 2).
 SQ SEQUENCE 88 AA; 9614 MW; 39445 CN;

Initial Score = 27 Optimized Score = 27 Significance = 17.85
 Residue Identity = 100% Matches = 27 Mismatches = 0
 Gaps = 0 Conservative Substitutions = 0

X 10 20
 KALPVVLENARILKNCVDAKMTEEDK
 |||||
 MLDAALPPCPTVAATADCEICPAVKRDVDFLTGTPDEYVEQVAQYKALPVVLENARILKNCVDAKMTEEDK
 10 20 30 40 X 50 60 70

X
 E
 |
 ENALSLLDKIYTSPLC
 X 80

3. US-08-300-510-2 (1-27)

POLG_FMDV1 GENOME POLYPROTEIN (NONSTRUCTURAL PROTEIN P20A; CO

ID POLG_FMDV1 STANDARD; PRT; 2333 AA.
 AC P03306;
 DT 21-JUL-1986 (REL. 01, CREATED)
 DT 21-JUL-1986 (REL. 01, LAST SEQUENCE UPDATE)
 DT 01-OCT-1994 (REL. 30, LAST ANNOTATION UPDATE)
 DE GENOME POLYPROTEIN (NONSTRUCTURAL PROTEIN P20A; COAT PROTEINS VP1 TO
 DE VP4; CORE PROTEIN P52; GENOME-LINKED PROTEINS VPG1 TO VPG3; PICORNAIN
 DE 3C (EC 3.4.22.28) (PROTEASE 3C) (P3C); RNA-DIRECTED RNA POLYMERASE
 DE (EC 2.7.7.48)).
 OS FOOT-AND-MOUTH DISEASE VIRUS (STRAIN A10-61) (APHTHOVIRUS A).
 OC VIRIDAE; SS-RNA NONENVELOPED VIRUSES; PICORNAVIRIDAE; APHTHOVIRUSES.
 RN [1]
 RP SEQUENCE FROM N.A.
 RM 84169547
 RA CARROLL A.R., ROWLANDS D.J., CLARKE B.E.;
 RL NUCLEIC ACIDS RES. 12:2461-2472(1984).
 RN [2]
 RP SEQUENCE OF 115-1048 FROM N.A.
 RM 82211814
 RA BOOTHROYD J.C., HARRIS T.J.R., ROWLANDS D.J., LOWE P.A.;
 RL GENE 17:153-161(1982).
 CC -!- PTM: SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
 CC -!- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAHEDRAL UNITS,
 CC EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,
 CC VP3, AND VP4.
 DR EMBL; X00429; PIFMDV1.
 DR PIR; A03908; GNMY2F.
 KW POLYPROTEIN; COAT PROTEIN; CORE PROTEIN; RNA-DIRECTED RNA POLYMERASE;
 KW HYDROLASE; THIOL PROTEASE; MYRISTYLATION.
 FT CHAIN 1 201 NONSTRUCTURAL PROTEIN P20A.
 FT CHAIN 202 286 COAT PROTEIN VP4.
 FT CHAIN 287 504 COAT PROTEIN VP2.
 FT CHAIN 505 725 COAT PROTEIN VP3.
 FT CHAIN 726 937 COAT PROTEIN VP1.
 FT CHAIN 938 1578 CORE PROTEIN P52.
 FT CHAIN 1579 1601 GENOME-LINKED PROTEIN VPG1.
 FT CHAIN 1602 1625 GENOME-LINKED PROTEIN VPG2.
 FT CHAIN 1626 1649 GENOME-LINKED PROTEIN VPG3.
 FT CHAIN 1650 1863 PROTEASE P20B.
 FT CHAIN 1864 2333 RNA-DIRECTED RNA POLYMERASE P56A.
 FT LIPID 202 202 MYRISTATE.
 FT CONFLICT 396 396 S -> C (IN REF. 2).
 FT CONFLICT 632 632 P -> L (IN REF. 2).

50 SEQUENCE 2333 AA: 239845 MW: 19388774 CN:

Initial Score = 13 Optimized Score = 14 Significance = 7.44
Residue Identity = 41% Matches = 15 Mismatches = 12
Gaps = 9 Conservative Substitutions = 0

```

GLFAYKAATRAGYCGGAVLAKDGADTFIVGTHSAGGNGVGYCSCVSRSLQKMKAHVDPEPHHEGLIVDTRD
1800      1810      1820      1830      1840      1850      1860      1870

              X              10      20      X
              KAL-----PVVLENARILKNCVDAKMTEEDKE
              ||              |||      |      |||||
VEERVHVMRKTKLAPTVA YGVFNPEFGPAALS NKDPRLNEGVVLD DVIFSKHKGD AKMTEEDKALFRRCAAD
1880      1890      1900      1910      1920      1930      X 1940

YASRLHSVLGTANAPLSIYEAIKGV DGLDAMEPDTAPGLPWALQ GKRRGALIDFENGTVGPEVEAALKLMEK
1950      1960      1970      1980      1990      2000      2010

REYKFACQTF LKDEIRPMEK
2020      2030

```

4. US-08-300-510-2 (1-27)

POLG_FMDVD GENOME POLYPROTEIN (NONSTRUCTURAL PROTEIN P20A; CO

```

ID  POLG_FMDVD      STANDARD;      PRT; 2332 AA.
AC  P03305;
DT  21-JUL-1986 (REL. 01, CREATED)
DT  21-JUL-1986 (REL. 01, LAST SEQUENCE UPDATE)
DT  01-OCT-1994 (REL. 30, LAST ANNOTATION UPDATE)
DE  GENOME POLYPROTEIN (NONSTRUCTURAL PROTEIN P20A; COAT PROTEINS VP1 TO
DE  VP4; CORE PROTEINS P12, P34, P14; GENOME-LINKED PROTEIN VPG; PROTEASE
DE  (EC 3.4.22.-); RNA-DIRECTED RNA POLYMERASE (EC 2.7.7.48)).
OS  FOOT-AND-MOUTH DISEASE VIRUS (STRAINS 01K AND 01BFS) (APHTHOVIRUS 0).
OC  VIRIDAE; SS-RNA NONENVELOPED VIRUSES; PICORNAVIRIDAE; APHTHOVIRUSES.
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=01K;
RM  84297249
RA  FORSS S., STREBEL K., BECK E., SCHALLER H.;
RL  NUCLEIC ACIDS RES. 12:6587-6601(1984).
RN  [2]
RP  SEQUENCE FROM N.A.
RC  STRAIN=01BFS;
RM  83143292
RA  MAKOFF A.J., PAYNTER C.A., ROWLANDS D.J., BOOTHROYD J.C.;
RL  NUCLEIC ACIDS RES. 10:8285-8295(1982).
RN  [3]
RP  X-RAY CRYSTALLOGRAPHY (2.9 ANGSTROMS).
RM  89143740
RA  ACHARYA R., FRY E., STUART D., FOX G., ROWLANDS D., BROWN F.;
RL  NATURE 337:709-716(1989).
CC  -!- THE STRAIN 01K SEQUENCE IS SHOWN.
CC  -!- PTM: SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
CC  -!- THE COAT PROTEIN VP1 CONTAINS THE MAIN ANTIGENIC DETERMINANTS OF
CC  THE VIRION; THEREFORE, CHANGES IN ITS SEQUENCE MUST BE
CC  RESPONSIBLE FOR THE HIGH ANTIGENIC VARIABILITY OF THE VIRUS.
CC  -!- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAEDRAL UNITS,
CC  EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,
CC  VP3, AND VP4.
DR  EMBL; X00871; PIFMDV2.
DR  EMBL; J02185; P101VP.
DR  PIR; A03907; GNNYF.
KW  POLYPROTEIN; COAT PROTEIN; CORE PROTEIN; RNA-DIRECTED RNA POLYMERASE;
KW  HYDROLASE; THIOL PROTEASE; MYRISTYLATION.
FT  CHAIN 1 201 NONSTRUCTURAL PROTEIN P20A.

```


FT	CHAIN	202	208	COAT PROTEIN VP4.
FT	CHAIN	287	504	COAT PROTEIN VP2.
FT	CHAIN	505	724	COAT PROTEIN VP3.
FT	CHAIN	725	937	COAT PROTEIN VP1.
FT	CHAIN	938	1107	CORE PROTEIN P12.
FT	CHAIN	1108	1425	CORE PROTEIN P34.
FT	CHAIN	1426	1578	CORE PROTEIN P14.
FT	CHAIN	1579	1649	GENOME-LINKED PROTEIN VPG.
FT	CHAIN	1650	1862	PROTEASE.
FT	CHAIN	1863	2332	RNA-DIRECTED RNA POLYMERASE.
FT	LIPID	202	202	MYRISTATE.
FT	DISULFID	511	511	INTERCHAIN (IN VP3 DIMER).
FT	DISULFID	406	858	IN VP2-VP1 DIMER.
FT	VARIANT	780	780	I -> V (IN STRAIN 01BFS).
FT	VARIANT	808	808	G -> R (IN STRAIN 01BFS).
FT	VARIANT	861	861	N -> S (IN STRAIN 01BFS).
SO	SEQUENCE	2332 AA;	258924 MW;	19411374 CN;

Initial Score = 11 Optimized Score = 12 Significance = 5.95
Residue Identity = 36% Matches = 13 Mismatches = 14
Gaps = 9 Conservative Substitutions = 0

GLFAYRAATKAGYCGGAVLAKDGDFTFIVGTHSAGGNGVGYCSCVSRSMLLKMKAHIDPEPHHEGLIVDTRD
1800 1810 1820 1830 1840 1850 1860 1870

X 10 20 X
KAL-----PVVLENARILKNCVDAKMTEEDKE
|| ||| | ||| |||
VEERVHVMRKTKLAPTVAHGTVFNPEFGPAALSNKDPRLNEGVLDEVIFSKHKGDTKMSEEDKALFRRCAAD
1880 1890 1900 1910 1920 1930 X 1940

YASRLHSVLGTANAPLSIYEAIKGVGDGLDAMEPDTAPGLPWALGGKRRGALIDFENGTVGPEVEAALKLMEK
1950 1960 1970 1980 1990 2000 2010

REYKFVCQTFLKDEIRPLEK
2020 2030

5. US-08-300-510-2 (1-27)

RPOL_KLULA PROBABLE DNA-DIRECTED RNA POLYMERASE (EC 2.7.7.6)

ID RPOL_KLULA STANDARD; PRT; 982 AA.
AC P05472;
DT 01-NOV-1988 (REL. 09, CREATED)
DT 01-NOV-1988 (REL. 09, LAST SEQUENCE UPDATE)
DT 01-NOV-1988 (REL. 09, LAST ANNOTATION UPDATE)
DE PROBABLE DNA-DIRECTED RNA POLYMERASE (EC 2.7.7.6) (KILLER PLASMID
DE PGKL2 PROTEIN 6).
OS KLUYVEROMYCES LACTIS (YEAST).
OG PLASMID PGKL-2.
OC EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOMYCETES.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CBS 2359;
RM 88289339
RA TOMMASINO S., RICCI S., GALEOTTI C.L.;
RL NUCLEIC ACIDS RES. 16:5863-5878(1988).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=IFO 1267;
RM 88335549
RA WILSON D.W., MEACOCK P.A.;
RL NUCLEIC ACIDS RES. 16:8097-8112(1988).
CC -!- FUNCTION: THE PRESENCE OF THE TWO LINEAR PLASMIDS, TERMED
CC PGKL1 AND PGKL2, IN STRAINS OF KLUYVEROMYCES LACTIS CONFERS
CC THE KILLER PHENOTYPE TO THE HOST CELL, BY PROMOTING THE

CC SECRETION OF A TOXIN ABLE TO INHIBIT THE GROWTH OF SENSITIVE
 CC STRAINS.
 DR EMBL; X07776; KLPGL2.
 DR EMBL; X07946; KLPOLK.
 DR PIR; S00964; S00964.
 KW TRANSCRIPTION; DNA-DIRECTED RNA POLYMERASE; PLASMID.
 FT CONFLICT 32 32 T -> N (IN REF. 2).
 FT CONFLICT 302 302 I -> K (IN REF. 2).
 FT CONFLICT 917 917 F -> C (IN REF. 2).
 SQ SEQUENCE 982 AA; 113961 MW; 4923764 CN;

Initial Score = 10 Optimized Score = 10 Significance = 5.21
 Residue Identity = 37% Matches = 10 Mismatches = 17
 Gaps = 0 Conservative Substitutions = 0

DILIGLGAWNTIKEIWSIDRSKIKIDSKTGRINWIRYDKEMEIGQYFKICLSYMRSLSGRDILIKNDKYSIVE
 640 650 660 670 680 690 700

X 10 20 X
 KALPVLLENARILKNCVDKMTTEEDKE
 || | | || | ||
 FDNSYLPKTDTMKFGDLVDLRIYKGIVMLPLCLRSTYLNKLYVDRKYSEAEKEVTKLLKSKNGAYHTLVE
 710 720 730 X 740 750 760 X 770

GHRVDR CIRSVIVPDP TLDIDTIKIPFGANICCEYGLLN RQPSLNVDSIKLVKLKQGSNKTIAINPLLCQSF
 780 790 800 810 820 830 840 850

NADFDGDEMNI
 860

6. US-08-300-510-2 (1-27)

POLG_FMDVA GENOME POLYPROTEIN (NONSTRUCTURAL PROTEIN P20A; CO

ID POLG_FMDVA STANDARD; PRT: 2332 AA.
 AC P03308; P03312;
 DT 21-JUL-1986 (REL. 01, CREATED)
 DT 01-JAN-1988 (REL. 06, LAST SEQUENCE UPDATE)
 DT 01-OCT-1994 (REL. 30, LAST ANNOTATION UPDATE)
 DE GENOME POLYPROTEIN (NONSTRUCTURAL PROTEIN P20A; COAT PROTEINS VP1 TO
 DE VP4; CORE PROTEINS X, P14, P41, P19; GENOME-LINKED PROTEINS VPG1 TO
 DE VPG3; PICORNAIN 3C (EC 3.4.22.28) (PROTEASE 3C) (P3C); RNA-DIRECTED
 DE RNA POLYMERASE (EC 2.7.7.48)).
 OS FOOT-AND-MOUTH DISEASE VIRUS (STRAIN A12) (APHTHOVIRUS A).
 OC VIRIDAE; SS-RNA NONENVELOPED VIRUSES; PICORNAVIRIDAE; APHTHOVIRUSES.
 RN [1]
 RP SEQUENCE FROM N.A.
 RM 85211015
 RA ROBERTSON B.H., GRUBMAN M.J., WEDDELL G.N., MOORE D.M., WELSH J.D.,
 RA FISCHER T., DOWBENKO D.J., YANSURA D.G., SMALL B., KLEID D.G.;
 RL J. VIROL. 54:651-660(1985).
 RN [2]
 RP SEQUENCE OF 1863-2332 FROM N.A.
 RM 83225613
 RA ROBERTSON B.H., MORGAN D.O., MOORE D.M., GRUBMAN M.J., CARD J.,
 RA FISCHER T., WEDDELL G.N., DOWBENKO D.J., YANSURA D.G.;
 RL VIROLOGY 126:614-623(1983).
 RN [3]
 RP SEQUENCE OF 715-955 FROM N.A.
 RM 82061853
 RA KLEID D.G., YANSURA D.G., SMALL B., DOWBENKO D.J., MOORE D.M.,
 RA GRUBMAN M.J., MCKERCHER P.D., MORGAN D.O., ROBERTSON B.H.,
 RA BACHRACH H.L.;
 RL SCIENCE 214:1125-1129(1981).
 CC -!- PTM: SPECIFIC ENZYMOLOGICAL CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
 CC -!- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAEDRAL UNITS,

CC EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,
 CC VP3, AND VP4.
 DR EMBL: M10975; APHA12CD.
 DR PIR: A25794; GNNY4F.
 KW POLYPROTEIN; COAT PROTEIN; CORE PROTEIN; RNA-DIRECTED RNA POLYMERASE;
 KW HYDROLASE; THIOL PROTEASE; MYRISTYLATION.
 FT CHAIN 1 200 NONSTRUCTURAL PROTEIN P20A.
 FT CHAIN 201 285 COAT PROTEIN VP4.
 FT CHAIN 286 503 COAT PROTEIN VP2.
 FT CHAIN 504 723 COAT PROTEIN VP3.
 FT CHAIN 724 937 COAT PROTEIN VP1.
 FT CHAIN 938 953 CORE PROTEIN X.
 FT CHAIN 954 1107 CORE PROTEIN P14.
 FT CHAIN 1108 1425 CORE PROTEIN P41.
 FT CHAIN 1426 1578 CORE PROTEIN P19.
 FT CHAIN 1579 1601 GENOME-LINKED PROTEIN VPG1.
 FT CHAIN 1602 1625 GENOME-LINKED PROTEIN VPG2.
 FT CHAIN 1626 1649 GENOME-LINKED PROTEIN VPG3.
 FT CHAIN 1650 1862 PROTEASE.
 FT CHAIN 1863 2332 RNA-DIRECTED RNA POLYMERASE.
 FT LIPID 201 201 MYRISTATE.
 SQ SEQUENCE 2332 AA; 259408 MW; 19347576 CN;

Initial Score = 10 Optimized Score = 11 Significance = 5.21
 Residue Identity = 33% Matches = 12 Mismatches = 15
 Gaps = 9 Conservative Substitutions = 0

SLFAYKAATKAGYCGGAVLAKDGADTFIVGTHSAGNGVGYCSCVSKSMLLRMKAHVDPEPQHEGLIVDTRD
 1800 1810 1820 1830 1840 1850 1860 1870
 X 10 20 X
 KAL-----PVVLENARILKNCVDAKMTEEDKE
 || ||| | ||| |||
 VEERVHVMRKTKLAPTVAHG VFNPEFGPAALSNKDPRLNEGVLDEVIFSKHKGDTKMSAEDKALFRACAAD
 1880 1890 1900 1910 1920 1930 X 1940
 YASRLHSVLGTANAPLSIYEAIKGV DGLDAMESDTAPGLP WAFQ GKRRGALIDFENGTVGPEVEAALKLMEK
 1950 1960 1970 1980 1990 2000 2010
 REYKFVCQTFLKDEIRPMEK
 2020 2030

7. US-08-300-510-2 (1-27)
 NHPA_YEAST NONHISTONE CHROMOSOMAL PROTEIN 6A.

ID NHPA_YEAST STANDARD; PRT; 93 AA.
 AC P11632;
 DT 01-OCT-1989 (REL. 12, CREATED)
 DT 01-OCT-1989 (REL. 12, LAST SEQUENCE UPDATE)
 DT 01-OCT-1993 (REL. 27, LAST ANNOTATION UPDATE)
 DE NONHISTONE CHROMOSOMAL PROTEIN 6A.
 GN NHP6A OR NHPA.
 OS SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).
 OC EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOMYCETES.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=DBY1091 / DKY1;
 RM 90153974
 RA KOLODRUBETZ D., BURGUM A.;
 RL J. BIOL. CHEM. 265:3234-3239(1990).
 RN [2]
 RP SEQUENCE FROM N.A.
 RM 93015901
 RA TERCERO J.C., RILES L.E., WICKNER R.B.;
 RL J. BIOL. CHEM. 267:20270-20276(1992).

CC -!- SIMILARITY: TO MAMMALIAN NONHISTONE PROTEIN HMG1.
 CC -!- SUBCELLULAR LOCATION: NUCLEAR.
 DR EMBL: X15317; SCNHP6A.
 DR EMBL: M95912; SCMAKNHP.
 DR PIR: A35072; A35072.
 DR PIR: C44031; C44031.
 KW NUCLEAR PROTEIN; CHROMOSOMAL PROTEIN; DNA-BINDING.
 FT DNA_BIND 18 93 HMG BOX.
 SQ SEQUENCE 93 AA; 10802 MW; 41127 CN;

Initial Score = 9 Optimized Score = 9 Significance = 4.46
 Residue Identity = 33% Matches = 9 Mismatches = 18
 Gaps = 0 Conservative Substitutions = 0

X 10
 KALPVVLENARIL
 |||
 MVTPREPKKRTRTRKKKDPNAPKRALSAYMFFANENRDIVRSENPDITFGQVGKKLGEKWKALTPEEKQPYEA
 10 20 30 40 50 60 70
 20 X
 KNCVDAKMTEEDKE
 | | | | |
 KAGADKKRYESEKELYNATLA
 80 X 90

8. US-08-300-510-2 (1-27)
 NHPB_YEAST NONHISTONE CHROMOSOMAL PROTEIN 6B.

ID NHPB_YEAST STANDARD; PRT: 99 AA.
 AC P11633;
 DT 01-OCT-1989 (REL. 12, CREATED)
 DT 01-OCT-1989 (REL. 12, LAST SEQUENCE UPDATE)
 DT 01-OCT-1993 (REL. 27, LAST ANNOTATION UPDATE)
 DE NONHISTONE CHROMOSOMAL PROTEIN 6B.
 GN NHP6B OR NHPB.
 OS SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).
 OC EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOMYCETES.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=DBY1091 / DKY1;
 RM 90153974
 RA KOLODRUBETZ D., BURGUM A.;
 RL J. BIOL. CHEM. 265:3234-3239(1990).
 CC -!- SIMILARITY: TO MAMMALIAN NONHISTONE PROTEIN HMG1.
 CC -!- SUBCELLULAR LOCATION: NUCLEAR.
 DR EMBL: X15318; SCNHP6B.
 DR PIR: B35072; B35072.
 KW NUCLEAR PROTEIN; CHROMOSOMAL PROTEIN; DNA-BINDING.
 FT DNA_BIND 24 99 HMG BOX.
 SQ SEQUENCE 99 AA; 11476 MW; 45060 CN;

Initial Score = 9 Optimized Score = 9 Significance = 4.46
 Residue Identity = 33% Matches = 9 Mismatches = 18
 Gaps = 0 Conservative Substitutions = 0

X
 KALPVVL
 |||
 MAATKEAKQPKPKKRTRRRKKDPNAPKRGLSAYMFFANENRDIVRSENPDVTFGQVGRILGERWKALTAEE
 10 20 30 40 50 60 X 70
 10 20 X
 ENARILKNCVDAKMTEEDKE
 | | | | |

9. US-08-300-510-2 (1-27)
HEMM_THEZO MYOHEMERYTHRIN.

ID HEMM_THEZO STANDARD; PRT: 118 AA.
AC P02247;
DT 21-JUL-1986 (REL. 01, CREATED)
DT 01-OCT-1993 (REL. 27, LAST SEQUENCE UPDATE)
DT 01-OCT-1993 (REL. 27, LAST ANNOTATION UPDATE)
DE MYOHEMERYTHRIN.
OS THEMISTE ZOSTERICOLA.
OC EUKARYOTA; METAZOA; SIPUNCULA; GOLFGINGIIDAE.
RN [1]
RP SEQUENCE.
RM 76136381
RA KLIPPENSTEIN G.L., COTE J.L., LUDLAM S.E.;
RL BIOCHEMISTRY 15:1128-1136(1976).
RN [2]
RP X-RAY CRYSTALLOGRAPHY (5.5 ANGSTROMS).
RM 75176901
RA HENDRICKSON W.A., KLIPPENSTEIN G.L., WARD K.B.;
RL PROC. NATL. ACAD. SCI. U.S.A. 72:2160-2164(1975).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (1.7 ANGSTROMS), AND REVISION TO 34-35.
RM 88062755
RA SHERIFF S., HENDRICKSON W.A., SMITH J.L.;
RL J. MOL. BIOL. 197:273-296(1987).
RN [4]
RP STRUCTURE.
RM 77165245
RA HENDRICKSON W.A., WARD K.B.;
RL J. BIOL. CHEM. 252:3012-3018(1977).
CC -!- FUNCTION: MYOHEMERYTHRIN IS AN OXYGEN-BINDING PROTEIN FOUND IN
CC THE RETRACTOR MUSCLES OF CERTAIN WORMS. THE OXYGEN-BINDING SITE
CC CONTAINS TWO IRON ATOMS.
CC -!- SUBUNIT: MONOMER.
CC -!- TISSUE SPECIFICITY: MUSCLE.
CC -!- SIMILARITY: TO HEMERYTHRINS FROM VARIOUS MARINE WORMS.
DR PIR: A37369; HRTHM.
DR PDB: 2MHR; 16-APR-88.
DR PROSITE: PS00550; HEMERYTHRINS.
KW OXYGEN TRANSPORT; MUSCLE PROTEIN; METAL-BINDING; IRON; 3D-STRUCTURE.
FT METAL 25 25 IRON 1.
FT METAL 54 54 IRON 1.
FT METAL 58 58 IRON 1 AND 2.
FT METAL 73 73 IRON 2.
FT METAL 77 77 IRON 2.
FT METAL 106 106 IRON 2.
FT METAL 111 111 IRON 1 AND 2.
FT HELIX 12 14
FT HELIX 19 37
FT HELIX 41 64
FT TURN 65 66
FT TURN 68 69
FT HELIX 70 85
FT TURN 86 86
FT HELIX 93 109
FT TURN 110 110
FT HELIX 111 114
FT TURN 115 117
SQ SEQUENCE 118 AA; 13778 MW; 75202 CN;

Initial Score = 9 Optimized Score = 11 Significance = 4.46

Residue Identity = 55% Matches = 12 Mismatches = 10
 Gaps = 9 Conservative Substitutions = 0

```

                                X
                                KALPVV-
                                |  ||
GWEIPEPYVWDESRVFEQLDEEHKKIFKGIFDCIRDNSAPNLATLVKVTNNHFTHEEAMMDAAKYSEVVP
  10      20      30      40      50      60      X 70

      10      20      X
-----LENARILKNCVDAKMTEEDKE
      ||  |  ||||  ||
HKKMHKDFLEKIGGLSAPVDAKNVDYCKEVLVNHKGTDFKYKGKL
  80      90      100      110

```

10. US-08-300-510-2 (1-27)

PHEA_FREDI C-PHYCOERYTHRIN ALPHA CHAIN.

ID PHEA_FREDI STANDARD; PRT; 164 AA.
 AC P05098;
 DT 13-AUG-1987 (REL. 05, CREATED)
 DT 13-AUG-1987 (REL. 05, LAST SEQUENCE UPDATE)
 DT 01-JUL-1993 (REL. 26, LAST ANNOTATION UPDATE)
 DE C-PHYCOERYTHRIN ALPHA CHAIN.
 GN CPEA.
 OS FREMYELLA DIPLOSIPHON (CALOTHRIX PCC 7601).
 OC PROKARYOTA; GRACILICUTES; OXYPHOTOBACTERIA;
 OC CYANOBACTERIA (BLUE-GREEN ALGAE); NOSTOCALES.
 RN [1]
 RP SEQUENCE FROM N.A.
 RM 87066711
 RA MAZEL D., GUGLIELMI G., HOUMARD J., SIDLER W., BRYANT D.A.,
 RA TANDEAU DE MARSAC N.;
 RL NUCLEIC ACIDS RES. 14:8279-8290(1986).
 RN [2]
 RP SEQUENCE.
 RM 87000169
 RA SIDLER W., KUMPF B., RUDIGER W., ZUBER H.;
 RL BIOL. CHEM. HOPPE-SEYLER 367:627-642(1986).
 CC -!- FUNCTION: LIGHT-HARVESTING PHOTOSYNTHETIC BILE PIGMENT-PROTEIN
 CC FROM THE PHYCOBILIPROTEIN COMPLEX.
 CC -!- SUBCELLULAR LOCATION: PERIPHERY OF THE RODS OF THE PHYCOBILISOME.
 CC -!- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA CHAIN.
 CC -!- PTM: CONTAINS ONE COVALENTLY LINKED BILIN CHROMOPHORE.
 CC -!- INDUCTION: IN RESPONSE TO GREEN LIGHT BUT NO TO RED LIGHT.
 DR EMBL; X04592; FDCPEAB.
 DR PIR; A25527; CFXCA.
 KW PHYCOBILISOME; ELECTRON TRANSPORT; PHOTOSYNTHESIS; BILE PIGMENT.
 FT BINDING 82 82 PHYCOERYTHROBILIN CHROMOPHORE.
 FT BINDING 139 139 PHYCOERYTHROBILIN CHROMOPHORE.
 SQ SEQUENCE 164 AA; 17626 MW; 123449 CN;

Initial Score = 9 Optimized Score = 10 Significance = 4.46
 Residue Identity = 29% Matches = 11 Mismatches = 16
 Gaps = 10 Conservative Substitutions = 0

```

      X      10      20      X
      KALPVV-----LENARILKNCVDAKHTTEEDKE
      |  |  ||  ||  ||  ||
MKSVVTTVIAAADAAGRFPSTSDLESVQGSIQRAAARLEAAEKLANNIDAVATEAYNACIKKYPYLNNSGEA
  10      20 X      30      40      50      60      70

NSTDTFKAKCARDIKHYLRLIQYSLVVGGTGPLDEWGIAGQREVYRALGLPTAPYVEALSFARNRGCAPRDM
  80      90      100      110      120      130      140

```

11. US-08-300-510-2 (1-27)
ATPF_PROMO ATP SYNTHASE B CHAIN, SODIUM ION SPECIFIC (EC 3.6.

ID ATPF_PROMO STANDARD; PRT; 168 AA.
AC P21904;
DT 01-MAY-1991 (REL. 18, CREATED)
DT 01-APR-1993 (REL. 25, LAST SEQUENCE UPDATE)
DT 01-OCT-1994 (REL. 30, LAST ANNOTATION UPDATE)
DE ATP SYNTHASE B CHAIN, SODIUM ION SPECIFIC (EC 3.6.1.37).
GN ATPF OR UNCF.
OS PROPIONIGENIUM MODESTUM.
OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; ANAEROBIC RODS;
OC BACTEROIDACEAE.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DSM 2376;
RM 91067471
RA KAIM G., LUDWIG W., DIMROTH P., SCHLEIFER K.H.;
RL NUCLEIC ACIDS RES. 18:6697-6697(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=DSM 2376;
RM 92339434
RA KAIM G., LUDWIG W., DIMROTH P., SCHLEIFER K.H.;
RL EUR. J. BIOCHEM. 207:463-470(1992).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=DSM 2376;
RM 91016937
RA ESSER U., KRUMHOLZ L.R., SIMONI R.D.;
RL NUCLEIC ACIDS RES. 18:5887-5888(1990).
RN [4]
RP SEQUENCE OF 1-7.
RM 93138123
RA GERIKE U., DIMROTH P.;
RL FEBS LETT. 316:89-92(1993).
CC -!- SUBUNIT: F-TYPE ATPASES HAVE 2 COMPONENTS, CF(1) - THE CATALYTIC
CC CORE - AND CF(0) - THE MEMBRANE PROTON CHANNEL. CF(1) HAS FIVE
CC SUBUNITS: ALPHA(3), BETA(3), GAMMA(1), DELTA(1), EPSILON(1). CF(0)
CC HAS THREE MAIN SUBUNITS: A, B AND C.
CC -!- THE ATPASE OF P.MODESTUM IS OF SPECIAL INTEREST BECAUSE IT
CC USES SODIUM IONS INSTEAD OF PROTONS AS THE PHYSIOLOGICAL
CC COUPLING ION.
CC -!- SIMILARITY: TO OTHER B SUBUNITS AND ALSO TO B' SUBUNITS.
DR EMBL; X54810; PMATPBS.
DR EMBL; X66102; PMATPACBD.
DR EMBL; X53960; PMUNC1.
DR EMBL; X58461; PMUNC2.
DR PIR; S12620; S12620.
DR PIR; S23323; S23323.
DR PIR; S23336; S23336.
KW SODIUM TRANSPORT; TRANSMEMBRANE; CF(0).
SQ SEQUENCE 168 AA; 19201 MW; 124854 CN;

Initial Score = 9 Optimized Score = 11 Significance = 4.46
Residue Identity = 44% Matches = 12 Mismatches = 14
Gaps = 1 Conservative Substitutions = 0

X 10 20 X
KALPVVLENARILKNCVDAKMTEEDKE
| || || | | | |||
MAPQNMPPAVSIDINMFQIINFLILMFFFFKKYFQKPIAKVL-DARKEKIANDLKQAEIDKEMAAKANGEAQQ

IVKSAKTEANEMLLRAEKKADERKETILKEANTQREKMLKSAEVEIEKMKEQARKELQLEVTDLAVKLAEKM
 80 90 100 110 120 130 140

INEKVDKIGANLLDQF
 150 160

12. US-08-300-510-2 (1-27)

UCRI_SYNP2 CYTOCHROME B6-F COMPLEX IRON-SULFUR SUBUNIT PRECUR

ID UCRI_SYNP2 STANDARD; PRT; 180 AA.
 AC P26292;
 DT 01-MAY-1992 (REL. 22, CREATED)
 DT 01-MAY-1992 (REL. 22, LAST SEQUENCE UPDATE)
 DT 01-MAY-1992 (REL. 22, LAST ANNOTATION UPDATE)
 DE CYTOCHROME B6-F COMPLEX IRON-SULFUR SUBUNIT PRECURSOR (EC 1.10.99.1)
 DE (RIESKE IRON-SULFUR PROTEIN).
 GN PETC.
 OS SYNECHOCOCCUS SP. (STRAIN PCC 7002) (AGMENELLUM QUADRUPLICATUM).
 OC PROKARYOTA; GRACILICUTES; OXYPHOTOBACTERIA;
 OC CYANOBACTERIA (BLUE-GREEN ALGAE); CHROOCOCCALES.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA WIDGER W.R.;
 RL SUBMITTED (XXX-1992) TO EMBL/GENBANK/DBJ DATA BANKS.
 CC -!- FUNCTION: COMPONENT OF THE CYTOCHROME B6/F COMPLEX WHICH IS PART
 CC OF THE CHLOROPLASTIC RESPIRATORY CHAIN.
 CC -!- THE RIESKE PROTEIN IS A HIGH POTENTIAL 2FE-2S PROTEIN.
 CC -!- CATALYTIC ACTIVITY: PLASTOQUINOL-1 + 2 OXYDIZED PLASTOCYANIN =
 CC PLASTOQUINONE + 2 REDUCED PLASTOCYANIN.
 CC -!- SUBUNIT: THE MAIN SUBUNITS OF COMPLEX B6-F ARE: CYTOCHROME B6,
 CC 17 KD POLYPEPTIDE (PETD), CYTOCHROME F AND THE RIESKE PROTEIN.
 CC -!- SIMILARITY: TO RIESKE PROTEINS FROM OTHER SOURCES (MITOCHONDRIA,
 CC BACTERIAL, CHLOROPLAST).
 DR EMBL; M74514; AQPETAC.
 DR PROSITE; PS00199; RIESKE_1.
 DR PROSITE; PS00200; RIESKE_2.
 KW ELECTRON TRANSPORT; INNER MEMBRANE; TRANSMEMBRANE; IRON-SULFUR;
 KW OXIDATIVE PHOSPHORYLATION; RESPIRATORY CHAIN.
 FT METAL 108 108 IRON-SULFUR CLUSTER (2FE-2S) (POTENTIAL).
 FT METAL 113 113 IRON-SULFUR CLUSTER (2FE-2S) (POTENTIAL).
 FT METAL 126 126 IRON-SULFUR CLUSTER (2FE-2S) (POTENTIAL).
 FT METAL 129 129 IRON-SULFUR CLUSTER (2FE-2S) (POTENTIAL).
 SQ SEQUENCE 180 AA; 19178 MW; 172619 CN;

Initial Score = 9 Optimized Score = 9 Significance = 4.46
 Residue Identity = 33% Matches = 9 Mismatches = 18
 Gaps = 0 Conservative Substitutions = 0

LYPVIKYFIPPSSGGAGGGVIAKDALGNDIIVSDYLQHTAGDRSLAQGLKGDPTYVVVEGDNTISSYGINA
 40 50 60 70 80 90 100

X 10 20 X
 KALPVVLENARILKNCVDAKMTEEDKE
 || | || ||
 ICTHLGCVVPWNTAENKFMCPCHGSQYDETCGVVRCAPLSLALVHAEVTEDDKISFTDWTETDFRTDEAPW
 110 120 130 X 140 150 160 170

WA
 180

13. US-08-300-510-2 (1-27)

XYNB_STRLI ENDO-1,4-BETA-XYLANASE B PRECURSOR (EC 3.2.1.8) (X

ID XYNB_STRLI STANDARD; PRT; 333 AA.
AC P26515;
DT 01-AUG-1992 (REL. 23, CREATED)
DT 01-AUG-1992 (REL. 23, LAST SEQUENCE UPDATE)
DT 01-JUN-1994 (REL. 29, LAST ANNOTATION UPDATE)
DE ENDO-1,4-BETA-XYLANASE B PRECURSOR (EC 3.2.1.8) (XYLANASE B)
DE (1,4-BETA-D-XYLAN XYLANOHYDROLASE B).
GN XLNB.
OS STREPTOMYCES LIVIDANS.
OC PROKARYOTA; FIRMICUTES; ACTINOMYCETALES; STREPTOMYCETACEAE.
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 41-71.
RC STRAIN=1326;
RM 92077439
RA SHARECK F., ROY C., YAGUCHI M., MOROSOLI R., KLUEPFEL D.;
RL GENE 107:75-82(1991).
CC -!- FUNCTION: CONTRIBUTES TO HYDROLYSE HEMICELLULOSE, THE MAJOR
CC COMPONENT OF PLANT CELL-WALLS. XLNA AND XLNB SEEM TO ACT
CC SEQUENTIALLY ON THE SUBSTRATE TO YIELD XYLOBIOSIDE AND XYLOSE
CC AS CARBON SOURCES.
CC -!- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-XYLOSIDIC
CC LINKAGES IN XYLANS.
CC -!- PATHWAY: XYLAN DEGRADATION.
CC -!- SUBCELLULAR LOCATION: EXTRACELLULAR.
CC -!- SIMILARITY: BELONGS TO CELLULASE FAMILY G (FAMILY 11 OF GLYCOSYL
CC HYDROLASES).
DR EMBL; M64552; SLXLNB.
DR PROSITE; PS00776; GLYCOSYL_HYDROL_F11_1.
DR PROSITE; PS00777; GLYCOSYL_HYDROL_F11_2.
KW XYLAN DEGRADATION; HYDROLASE; GLYCOSIDASE; SIGNAL.
FT SIGNAL 1 40
FT CHAIN 41 333 XYLANASE B.
FT ACT_SITE 127 127 BY SIMILARITY.
FT ACT_SITE 194 194 BY SIMILARITY.
FT ACT_SITE 217 217 BY SIMILARITY.
SQ SEQUENCE 333 AA; 35426 MW; 558782 CN;

Initial Score = 9 Optimized Score = 9 Significance = 4.46
Residue Identity = 33% Matches = 9 Mismatches = 18
Gaps = 0 Conservative Substitutions = 0

RTGGTITTTGNHFDARAGMPLGNFSYYMIMATEGYQSSGTSSINVGGTGGGDSGGGDNGGGGGGCTRRCP
190 200 210 220 230 240 250

X 10 20 X
KALPVVLENARILKNCVDAKMTEEDKE

GRSGATGTTSTSPSAAPRLDGDDEAVPGEGPVDLERQRLSQCADADRQLNGSGNNWGATIGANANWTWPS
260 270 280 X 290 300 310 320

VSCSAG
330

14. US-08-300-510-2 (1-27)
YB09_YEAST HYPOTHETICAL 38.7 KD PROTEIN IN RPB5-CDC28 INTERGE

ID YB09_YEAST STANDARD; PRT; 347 AA.
AC P38286;
DT 01-OCT-1994 (REL. 30, CREATED)
DT 01-OCT-1994 (REL. 30, LAST SEQUENCE UPDATE)
DT 01-OCT-1994 (REL. 30, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 38.7 KD PROTEIN IN RPB5-CDC28 INTERGENIC REGION.
GN YBR159W OR YBR1209.
OS SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).

OC EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOMYCETES.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288C;
 RA ENTIAN K.D., KOETTER P., ROSE M., BECKER J., GREY M., LI Z.,
 RA NIEGEMANN E., SCHENK-GROENINGER R., SERVOS J., WEHNER E.,
 RA WOLTER R., BRENDDEL M., BAUER J., BRAUN H., DERN K., DUESTERHUS S.,
 RA GRUENBEIN R., HEDGES D., KIESAU P., KOROL S., KREMS B., PROFT M.,
 RA SIEGERS K., BAUR A., BOLES E., MIDSGA T.,
 RA SCHAAFF-GERSTENSCHLAEGER I., ZIMMERMANN F.K.;
 RL SUBMITTED (AUG-1994) TO EMBL/GENBANK/DBJ DATA BANKS.
 DR EMBL: Z36028; SCYBR159W.
 DR PIR: S46030; S46030.
 KW HYPOTHETICAL PROTEIN; TRANSMEMBRANE.
 FT TRANSMEM 17 37 POTENTIAL.
 FT TRANSMEM 39 59 POTENTIAL.
 SQ SEQUENCE 347 AA; 38708 MW; 656470 CN;

Initial Score = 9 Optimized Score = 9 Significance = 4.46
 Residue Identity = 33% Matches = 9 Mismatches = 18
 Gaps = 0 Conservative Substitutions = 0

MTFMQQLQEAGERFRNCINGLLWVVFGLGVLKCTTSLRFLALIFDLFLLPAVNFDKYGAKTGKYCAITGASD
 10 20 30 40 50 60 70
 X 10 20 X
 KALPVVLENARILKNCVDAKMTEEDKE
 || || | |||
 GIGKEFARQMAKRGFNLVLISRTQSKLEALQKELEDQHHVVVKILADIAEDKESNYESIKELCAQLPITVL
 80 90 100 110 120 X 130 140
 VNNVGGSHSIPVPFLETEEKELRNIITINNTATLLITQIIAPKIVETVKAENKKSGTRGLILTMGSFGGLIP
 150 160 170 180 190 200 210
 TPLLATYSGS
 220

15. US-08-300-510-2 (1-27)
 ADH4_YEAST ALCOHOL DEHYDROGENASE IV (EC 1.1.1.1).

ID ADH4_YEAST STANDARD; PRT: 382 AA.
 AC P10127;
 DT 01-MAR-1989 (REL. 10, CREATED)
 DT 01-MAR-1989 (REL. 10, LAST SEQUENCE UPDATE)
 DT 01-OCT-1993 (REL. 27, LAST ANNOTATION UPDATE)
 DE ALCOHOL DEHYDROGENASE IV (EC 1.1.1.1).
 GN ADH4.
 OS SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).
 OC EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOMYCETES.
 RN [1]
 RP SEQUENCE FROM N.A.
 RM 88038383
 RA WILLIAMSON V.M., PAQUIN C.E.;
 RL MOL. GEN. GENET. 209:374-381(1987).
 CC -!- FUNCTION: NOT KNOWN YET, AS IN YEAST ADH4 IS NOT EXPRESSED UNDER
 CC LABORATORY CONDITIONS EXCEPT UPON INSERTION OF A TY AT THE ADH4
 CC LOCUS OR AMPLIFICATION OF ADH4.
 CC -!- CATALYTIC ACTIVITY: ALCOHOL + NAD(+) = ALDEHYDE OR KETONE + NADH.
 CC -!- SIMILARITY: BELONGS TO THE IRON-CONTAINING ALCOHOL DEHYDROGENASE
 CC FAMILY.
 DR EMBL: X05992; SCADH4.
 DR PIR: S07614; DEBY4.
 DR PROSITE: PS00060; ADH_IRON_2.
 DR PROSITE: PS00913; ADH_IRON_1.
 KW OXIDOREDUCTASE; NAD; IRON.

Seq SEQUENCE 382 AA: 41013 MW: 872276 Da

Initial Score = 9 Optimized Score = 9 Significance = 4.46
Residue Identity = 33% Matches = 9 Mismatches = 18
Gaps = 0 Conservative Substitutions = 0

DLINESLVAAYKDGKDKKARTDMCYAEYLAGMAFNNASLGYVHALAHQLGGFYHLPHGVCNAVLLPHVQEAN
230 240 250 260 270 280 290

X 10 20 X
KALPVVLENARILKNCVDAKMTEEDKE

MQCPKAKKRLGEIALHCGASQEDPEETIKALHVLNRTMNIPRNKDLGVKTEDFDILAEGHAMHDACHLTNPV
300 310 320 X 330 340 350 360

QFTKEQVVAIKKAVEY
370 380

> O <
O| |O IntelliGenetics
> O <

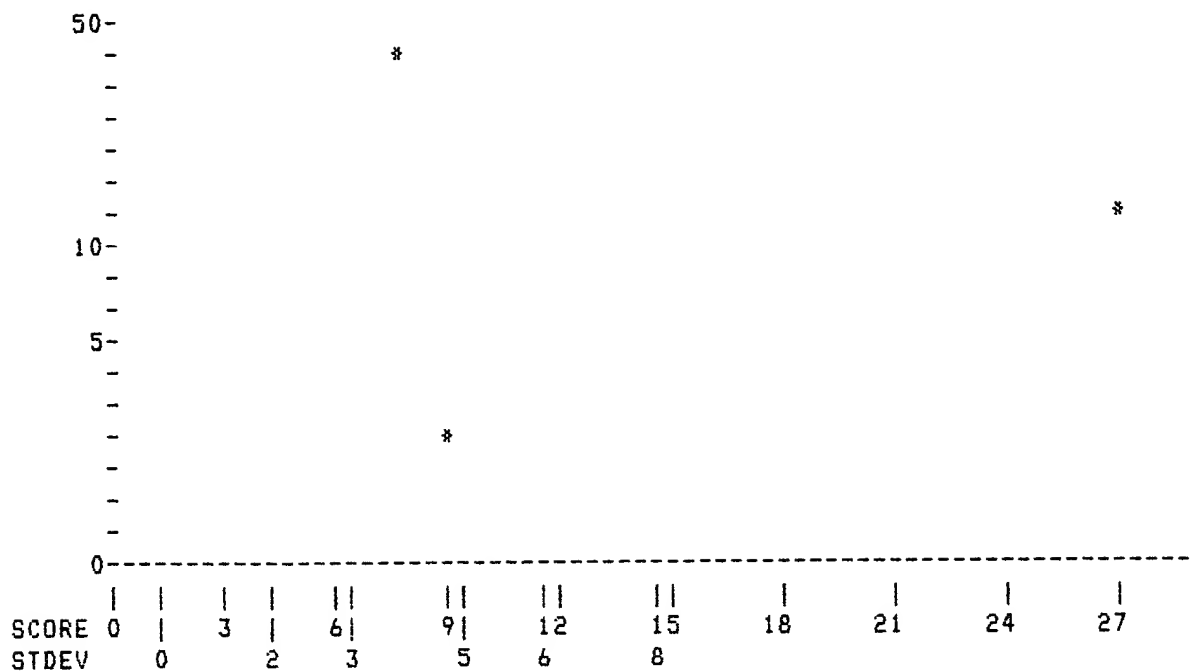
FastDB - Fast Pairwise Comparison of Sequences
Release 5.4

Results file 2ppat.res made by on Fri 24 Mar 95 7:56:28-PST.

Query sequence being compared: US-08-300-510-2 (1-27)
Number of sequences searched: 50375
Number of scores above cutoff: 4007

Results of the initial comparison of US-08-300-510-2 (1-27) with:
Data bank : A-GeneSeq 17, all entries

100000-
-
N -
U50000-
M -
B -
E - *
R -
*
Q -
F10000-
-
S - * *
E 5000-
Q -
U -
E - *
N -
C -
E -
S 1000- *
-
*
500-
-
-
-
*
100-
-



PARAMETERS

Similarity matrix	Unitary	K-tuple	2
Mismatch penalty	1	Joining penalty	20
Gap penalty	1.00	Window size	27
Gap size penalty	0.05		
Cutoff score	0		
Randomization group	0		
Initial scores to save	40	Alignments to save	15
Optimized scores to save	0	Display context	100

SEARCH STATISTICS

Scores:	Mean	Median	Standard Deviation
	2	3	1.64
Times:	CPU	Total Elapsed	
	00:00:21.01	00:00:22.00	
Number of residues:	6065180		
Number of sequences searched:	50375		
Number of scores above cutoff:	4007		

Cut-off raised to 2.
 Cut-off raised to 3.
 Cut-off raised to 4.
 Cut-off raised to 5.
 Cut-off raised to 6.

The scores below are sorted by initial score.
 Significance is calculated based on initial score.

2 100% identical sequences to the query sequence were found:

Sequence Name	Description	Length	Init. Score	Opt. Score	Sig. Frame
1. R41976	Human T cell reactive feline	27	27	27	15.21 0
2. R36543	Peptide Y.	27	27	27	15.21 0

9 100% similar sequences to the query sequence were found:

Sequence Name	Description	Init. Opt.		Length	Score	Score	Sig.	Frame
3. R12120	TRFP chain 1 with leader B.	96	27	27	15.21	0		
4. R27368	TRFP Chain #1 with C1 leader	96	27	27	15.21	0		
5. R36548	Recombitope YZX.	96	27	27	15.21	0		
6. R27367	TRFP Chain #1 with C1 leader	94	27	27	15.21	0		
7. R12119	TRFP chain 1 with leader A.	94	27	27	15.21	0		
8. R36539	TRFP chain 1 (with Leader A).	92	27	27	15.21	0		
9. R41983	Human T cell reactive feline	92	27	27	15.21	0		
10. R36540	TRFP chain 1 (with Leader B).	88	27	27	15.21	0		
11. R41984	Human T cell reactive feline	88	27	27	15.21	0		

The list of other best scores is:

		Init. Opt.				
Sequence Name	Description	Length	Score	Score	Sig.	Frame

	**** 4 standard deviations above mean ****					
12. R13139	B.burgdorferi strain PKo p100	663	9	9	4.26	0
13. R44489	Sequence of all or part of a	1429	9	9	4.26	0
	**** 3 standard deviations above mean ****					
14. R15140	Vascular injury affinity pept	18	8	8	3.65	0
15. R15127	Vascular injury affinity pept	18	8	8	3.65	0
16. R38912	Recombitope XYZ.	35	8	12	3.65	0
17. P30230	Sequence of interferon IFN- α 1	189	8	9	3.65	0
18. R07678	IFN- α 61.	189	8	9	3.65	0
19. R47338	Peptide fragment of tetracycl	194	8	9	3.65	0
20. R55342	Sequence of rabbit α -1 ac	201	8	9	3.65	0
21. R37299	Plant type I RIP Saporin 6.	259	8	9	3.65	0
22. R43955	Saporin from clone M13 mp18-G	268	8	9	3.65	0
23. R43954	Saporin from clone M13 mp18-G	268	8	9	3.65	0
24. R43953	Saporin from clone M13 mp18-G	268	8	9	3.65	0
25. R43952	Saporin from clone M13 mp18-G	268	8	9	3.65	0
26. R43951	Saporin from clone M13 mp18-G	268	8	9	3.65	0
27. R26995	Human IGFBP-5.	272	8	10	3.65	0
28. P90957	Ribosomal inactivating protei	280	8	9	3.65	0
29. R51233	Heat resistant alkali proteas	361	8	9	3.65	0
30. R49248	Actin.	375	8	9	3.65	0
31. R22096	Actin.	375	8	9	3.65	0
32. R22026	A. chrysogenum actin.	375	8	9	3.65	0
33. R25276	SCC antigen.	390	8	10	3.65	0
34. R43958	Saporin/FGF fusion protein.	410	8	9	3.65	0
35. R43957	Saporin/FGF fusion protein.	410	8	9	3.65	0
36. P60230	Dihydroxyacetone-synthetase.	702	8	9	3.65	0
37. R11607	Recombinant dihydroxyacetone	702	8	9	3.65	0
38. R49507	Human LIF-R clone 65.	719	8	9	3.65	0
39. R25069	mLIF-R.	719	8	9	3.65	0
40. R52027	Protein with Oxetanocin-A pro	744	8	8	3.65	0

1. US-08-300-510-2 (1-27)

R41976 Human T cell reactive feline protein fragment Y.

ID R41976 standard; peptide; 27 AA.
AC R41976;
DT 21-APR-1994 (first entry)
DE Human T cell reactive feline protein fragment Y.
KW Human; T cell; reactive; feline; protein; immune response; antigen;
KW tolerance; mammal; Dermatophagoides; Felis; Ambrosia; Lolium; Canis;
KW Cryptomeria; Alternaria; Alder; Betula; Quercus; Olea; Artemesia;
KW Plantago; Parietaria; Blattella; Apis; Periplaneta; autoantigen; ss.
OS Homo sapiens.
PN W09319178-A.
PD 30-SEP-1993.

PR 25-MAR-1992; US-857311.
 PR 15-MAY-1992; US-884718.
 PR 15-JAN-1993; US-006116.
 PA (IMMU-) IMMUNOLOGIC PHARM CORP.
 PI Briner TJ, Garman RD, Gefter ML, Greenstein JL;
 PI Kuo M, Morville M;
 DR WPI; 93-320744/40.
 PT New peptide(s) for inducing tolerance - comprise one or more
 PT epitope(s) of an allergen administered subcutaneously, for
 PT treating sensitivity to cats, bees, etc.
 PS Claim 1; Fig 3; 107pp; English.
 CC The sequences given in R41975-82 are peptides derived from a human T
 CC cell reactive feline protein. These peptides are used in a
 CC therapeutic composition which is useful in treating diseases which
 CC involve an immune response to a protein antigen. This composition
 CC may be used to induce tolerance in a mammal to Dermatophagoides,
 CC Felis, Ambrosia, Lolium, Cryptomeria, Alternaria, Alder, Betula,
 CC Quercus, Olea, Artemesia, Plantago, Parietaria, Canis, Blattella,
 CC Apis, Periplaneta and to autoantigens in humans.
 SQ Sequence 27 AA;
 SQ 3 A; 1 R; 2 N; 2 D; 0 B; 1 C; 0 Q; 4 E; 0 Z; 0 G; 0 H;
 SQ 1 I; 3 L; 4 K; 1 M; 0 F; 1 P; 0 S; 1 T; 0 W; 0 Y; 3 V;

Initial Score = 27 Optimized Score = 27 Significance = 15.21
 Residue Identity = 100% Matches = 27 Mismatches = 0
 Gaps = 0 Conservative Substitutions = 0

```

      X      10      20      X
      KALPVLLENARILKNCVDAKMTTEEDKE
      |||||
      KALPVLLENARILKNCVDAKMTTEEDKE
      X      10      20      X
  
```

2. US-08-300-510-2 (1-27)
 R36543 Peptide Y.

ID R36543 standard; Protein; 27 AA.
 AC R36543;
 DT 12-AUG-1993 (first entry)
 DE Peptide Y.
 KW Human T cell reactive feline protein; TRFP; epitope; recombitope.
 OS Felis.
 PN W09308280-A.
 PD 29-APR-1993.
 PF 16-OCT-1992; U08694.
 PR 16-OCT-1991; US-777859.
 PR 13-DEC-1991; US-807529.
 PA (IMMU-) IMMULOGIC PHARM CORP.
 PI Bond JF, Garman RD, Kuo M, Morgenstern JP, Morville M;
 PI Rogers BL;
 DR WPI; 93-152473/18.
 PT Recombitope peptide having T-cell stimulating activity - for the
 PT diagnosis and treatment of sensitivity to protein allergens,
 PT auto:antigens and protein antigens
 PS Disclosure; Fig 4; 73pp; English.
 CC Chains 1 and 2 of the TRFP have been recombinantly expressed in E.
 CC coli and purified. T cell epitope studies using overlapping peptide
 CC regions derived from the TRFP amino acids sequence were used to
 CC identify multiple T cell epitopes in each chain of TRFP. DNA
 CC constructs were assembled in which 3 regions (encoding peptides X,
 CC Y and Z) were linked to produce DNA constructs encoding recombitope-
 CC peptides.
 SQ Sequence 27 AA;
 SQ 3 A; 1 R; 2 N; 2 D; 0 B; 1 C; 0 Q; 4 E; 0 Z; 0 G; 0 H;

50 1 17 3 L: 4 K: 1 M: 0 P: 1 P: 0 S: 1 T: 0 W: 0 Y: 3 V:

Initial Score = 27 Optimized Score = 27 Significance = 15.21
Residue Identity = 100% Matches = 27 Mismatches = 0
Gaps = 0 Conservative Substitutions = 0

```

X      10      20      X
KALPVVLENARILKNCVDAKMTEEDKE
|||||
KALPVVLENARILKNCVDAKMTEEDKE
X      10      20      X
```

3. US-08-300-510-2 (1-27)

R12120 TRFP chain 1 with leader B.

ID R12120 standard; Protein; 96 AA.
AC R12120;
DT 26-JUL-1991 (first entry)
DE TRFP chain 1 with leader B.
KW Human T cell reactive feline protein; cat allergens.
OS Felis catus.
FH Key Location/Qualifiers
FT Peptide 9..26
FT /label= Leader B
FT Protein 27..96
FT /label= TRFP Chain 1
PN W09106571-A.
PD 16-MAY-1991.
PF 02-NOV-1990; U06548.
PR 03-NOV-1989; US-431565.
PA (IMMU-) IMMULOGIC PHARM COR.
PI Geffer ML, Garman RD, Greenstein JL, Juo M, Rogers BL;
PI Brauer AW;
DR WPI; 91-164136/22.
DR N-PSDB; Q11837.
PT New pure covalently linked human T cell reactive feline protein -
PT and modified peptide(s), used to reduce effects of cat allergens
PT and to diagnose sensitivity to allergens.
PS Claim 2; Fig 1; 70pp; English.
CC Poly-A mRNA from cat parotid and mandibular glands was used to
CC produce cDNA clones for both chain 1 and chain 2 of TRFP. These
CC clones were then used to screen a cat genomic library. Chain 1
CC exists in two forms having different leader sequences (A and B).
CC The sequence can be used to express the protein and peptide derivs.
CC which stimulate T-cells in persons allergic to cats. The peptides
CC can be used to reduce/eliminate the allergic response partic. by
CC modificn. of lymphokine prodn. by the T-cells. They can also be
CC used to identify epitopes responsible for sensitivity. The DNA can
CC be used to detect comparable sequence in other species, and also
CC for prodn. of modified forms of TRFP esp. showing reduced binding
CC to IgE and thus reduced tendency to cause adverse reactions.
CC See also R12119-R12123.
SQ Sequence 96 AA;
SQ 12 A; 4 R; 3 N; 8 D; 0 B; 6 C; 2 Q; 7 E; 0 Z; 1 G; 0 H;
SQ 3 I; 11 L; 7 K; 2 M; 1 F; 7 P; 3 S; 6 T; 2 W; 3 Y; 8 V;

Initial Score = 27 Optimized Score = 27 Significance = 15.21
Residue Identity = 100% Matches = 27 Mismatches = 0
Gaps = 0 Conservative Substitutions = 0

```

X      10
KALPVVLENARILKNCVD
|||||
AWRC5WKRM LDAALPPCPTVAATADCEICPAVKRDVDLFTGTPDEYVEQVAQYKALPVVLENARILKNCVD
10      20      30      40      50      X      60      70
```

```

20      X
AKMTEEDKE
|||||||
AKMTEEDKENALSLLDKIYTSPLC
      80      90

```

4. US-08-300-510-2 (1-27)

R27368 TRFP Chain #1 with C1 leader B sequence.

```

ID   R27368 standard; protein; 96 AA.
AC   R27368;
DT   25-FEB-1993 (first entry)
DE   TRFP Chain #1 with C1 leader B sequence.
KW   T cell reactive feline protein; cat allergy; allergic; IgE;
KW   desensitizing;
OS   Felis domesticus.
FH   Key          Location/Qualifiers
FT   Peptide      1..27
FT   /label= Leader B
FT   Protein      28..96
FT   /label= TRFP chain #1
PN   W09215613-A.
PD   17-SEP-1992.
PF   20-FEB-1992; U01344.
PR   28-FEB-1991; US-662193.
PA   (IMMU-) IMMULOGIC PHARM CORP.
PI   Bond J, Kuo M;
DR   WPI; 92-331670/40.
PT   Modified human T-cell reactive feline protein - stimulates T-cell
PT   in individuals allergic to cats and shows reduced
PT   histamine-releasing properties
PS   Claim 1; Fig 1; 35pp; English.
CC   This sequence represents a modified human T-cell reactive feline
CC   protein which stimulates T-cells from an individual who is allergic
CC   to cats, but which interacts with human IgE to a lesser extent than
CC   does affinity purified TRFP. The protein is modified by treating
CC   with either a mild alkali (pH 12.5-13.5 , KOH, NaOH, LiOH or tertiary
CC   amines) or an enzyme which removes O-linked groups (carbohydrate
CC   moieties). It is useful in desensitising people who are allergic to cats.
SQ   Sequence 96 AA;
SQ   12 A; 4 R; 3 N; 8 D; 1 B; 6 C; 2 Q; 7 E; 0 Z; 1 G; 0 H;
SQ   3 I; 11 L; 7 K; 2 M; 1 F; 7 P; 3 S; 6 T; 2 W; 3 Y; 7 V;

```

```

Initial Score      =      27   Optimized Score =      27   Significance = 15.21
Residue Identity   =    100%   Matches          =      27   Mismatches   =      0
Gaps               =          0   Conservative Substitutions =      0

```

```

                                     X      10
                                     KALPVVLENARILKNCVD
                                     |||||||||||||||
AWRCSWKRLDAALPPCPTBAATADCEICPAVKRDVDLFLTGTPDEYVEQVAQYKALPVVLENARILKNCVD
      10      20      30      40      50      X      60      70

```

```

20      X
AKMTEEDKE
|||||||
AKMTEEDKENALSLLDKIYTSPLC
      80      90

```

5. US-08-300-510-2 (1-27)

R36548 Recombitope YZX.

ID R36548 standard; Protein; 96 AA.

AC R38048;
 DT 12-AUG-1993 (first entry)
 DE Recombitope YZX.
 KW Human T cell reactive feline protein; TRFP; epitope; recombiteope
 KW sensitivity; Felis domesticus.
 OS Synthetic.
 FH Key Location/Qualifiers
 FT Cleavage_site 14..15
 FT /label= thrombin_cleavage_site
 PN W09308280-A.
 PD 29-APR-1993.
 PF 16-OCT-1992; U08694.
 PR 16-OCT-1991; US-777859.
 PR 13-DEC-1991; US-807529.
 PA (IMMU-) IMMULOGIC PHARM CORP.
 PI Bond JF, Garman RD, Kuo M, Morgenstern JP, Morville M;
 PI Rogers BL;
 DR WPI; 93-152473/18.
 DR N-PSDB; Q41572.
 PT Recombitope peptide having T-cell stimulating activity - for the
 PT diagnosis and treatment of sensitivity to protein allergens,
 PT auto:antigens and protein antigens
 PS Disclosure; Fig 8; 73pp; English.
 CC Preferred recombiteope peptides for treating sensitivity to Felis
 CC domesticus are derived from the the genus Felis and comprise
 CC regions selected from peptides X, Y, Z, A and B, of TRFP, and
 CC modifications thereof, such as peptide C.
 CC Oligonucleotides C, D, E, F, G, H and I are used in the
 CC construction of recombiteope peptide YZX.
 SQ Sequence 96 AA:
 SQ 8 A; 4 R; 5 N; 6 D; 0 B; 1 C; 2 Q; 10 E; 0 Z; 4 G; 6 H;
 SQ 1 I; 12 L; 7 K; 2 M; 4 F; 5 P; 2 S; 5 T; 0 W; 2 Y; 10 V;

Initial Score = 27 Optimized Score = 27 Significance = 15.21
 Residue Identity = 100% Matches = 27 Mismatches = 0
 Gaps = 0 Conservative Substitutions = 0

X 10 20 X
 KALPVVLENARILKNCVDAKMTEEDKE
 ||||||||||||||||||||
 MGHHHHHHEFLVPRGSKALPVVLENARILKNCVDAKMTEEDKEFFAVANGNELLLDLSLTQVNA TE PERKRD
 10 X 20 30 40 X 50 60 70
 VDLFLTGTPEYVEQVAQYKALPV
 80 90

6. US-08-300-510-2 (1-27)
 R27367 TRFP Chain #1 with C1 leader A sequence.

ID R27367 standard; protein; 94 AA.
 AC R27367;
 DT 25-FEB-1993 (first entry)
 DE TRFP Chain #1 with C1 leader A sequence.
 KW T cell reactive feline protein.
 OS Felis domesticus.
 FH Key Location/Qualifiers
 FT Peptide 1..25
 FT /label= Leader A
 FT Protein 25..94
 FT /label= TRFP chain #1
 PN W09215613-A.
 PD 17-SEP-1992.
 PF 20-FEB-1992; U01344.
 PR 28-FEB-1991; US-662193.
 PA (IMMU-) IMMULOGIC PHARM CORP.

DR WPI: 92-331670/40.
 PT Modified human T-cell reactive feline protein - stimulates T-cell
 PT in individuals allergic to cats and shows reduced
 PT histamine-releasing properties
 PS Claim 1; Fig 1; 35pp; English.
 CC This sequence represents a modified human T-cell reactive feline
 CC protein which stimulates T-cells from an individual who is allergic
 CC to cats, but which interacts with human IgE to a lesser extent than
 CC does affinity purified TRFP. The protein is modified by treating
 CC with either a mild alkali (pH 12.5-13.5, KOH, NaOH, LiOH or tertiary
 CC amines) or an enzyme which removes O-linked groups (carbohydrate
 CC moieties). It is useful in desensitising people who are allergic to cats.
 SQ Sequence 94 AA;
 SQ 9 A: 3 R: 4 N: 6 D: 0 B: 5 C: 2 Q: 7 E: 0 Z: 4 G: 0 H:
 SQ 5 I: 15 L: 7 K: 2 M: 1 F: 4 P: 2 S: 4 T: 2 W: 3 Y: 9 V:

Initial Score = 27 Optimized Score = 27 Significance = 15.21
 Residue Identity = 100% Matches = 27 Mismatches = 0
 Gaps = 0 Conservative Substitutions = 0

```

                                X      10      20
                                KALPVVLENARILKNCVDAK
                                |||||
CIMKGARVLVLLWAALLLIWGGNCEICPAVKRDVDLFLTGTPDEYVEQVAQYKALPVVLENARILKNCVDAK
      10      20      30      40      50 X      60      70

      X
MTEEDKE
      |||||
MTEEDKENALSLLDKIYTSPLC
      80      90
  
```

7. US-08-300-510-2 (1-27)

R12119 TRFP chain 1 with leader A.

ID R12119 standard; Protein: 94 AA.
 AC R12119;
 DT 26-JUL-1991 (first entry)
 DE TRFP chain 1 with leader A.
 KW Human T cell reactive feline protein; cat allergens.
 OS Felis catus.
 FH Key Location/Qualifiers
 FT Peptide 3..24
 FT /label= Leader B
 FT Protein 25..94
 FT /label= TRFP Chain 1
 PN W09106571-A.
 PD 16-MAY-1991.
 PF 02-NOV-1990; U06548.
 PR 03-NOV-1989; US-431565.
 PA (IMMU-) IMMULOGIC PHARM COR.
 PI Gefter ML, Garman RD, Greenstein JL, Juo M, Rogers BL;
 PI Brauer AW;
 DR WPI: 91-164136/22.
 DR N-PSDB; Q11836.
 PT New pure covalently linked human T cell reactive feline protein -
 PT and modified peptide(s), used to reduce effects of cat allergens
 PT and to diagnose sensitivity to allergens.
 PS Claim 2; Fig 1; 70pp; English.
 CC Poly-A mRNA from cat parotid and mandibular glands was used to
 CC produce cDNA clones for both chain 1 and chain 2 of TRFP. These
 CC clones were then used to screen a cat genomic library. Chain 1
 CC exists in two forms having different leader sequences (A and B).
 CC The sequence can be used to express the protein and peptide derivs.

CC which stimulate T-cells in persons allergic to cats. The peptides
 CC can be used to reduce/eliminate the allergic response partic. by
 CC modificn. of lymphokine prodn. by the T-cells. They can also be
 CC used to identify epitopes responsible for sensitivity. The DNA can
 CC be used to detect comparable sequence in other species, and also
 CC for prodn. of modified forms of TRFP esp. showing reduced binding
 CC to IgE and thus reduced tendency to cause adverse reactions.
 CC See also R12120-R12123.

SQ Sequence 94 AA;
 SQ 9 A; 3 R; 4 N; 6 D; 0 B; 5 C; 2 Q; 7 E; 0 Z; 4 G; 0 H;
 SQ 5 I; 15 L; 7 K; 2 M; 1 F; 4 P; 2 S; 4 T; 2 W; 3 Y; 9 V;

Initial Score = 27 Optimized Score = 27 Significance = 15.21
 Residue Identity = 100% Matches = 27 Mismatches = 0
 Gaps = 0 Conservative Substitutions = 0

X 10 20
 KALPVVLENARILKNCVDAK
 |||||
 CIMKGARVLVLLWAALLLIWGNCICPAVKRDVDLFTGTPDEYVEQVAQYKALPVVLENARILKNCVDAK
 10 20 30 40 50 X 60 70

X
 MTEEDKE
 |||||
 MTEEDKENALSLLDKIYTSPLC
 80 90

8. US-08-300-510-2 (1-27)
 R36539 TRFP chain 1 (with Leader A).

ID R36539 standard; Protein; 92 AA.
 AC R36539;
 DT 12-AUG-1993 (first entry)
 DE TRFP chain 1 (with Leader A).
 KW Human T cell reactive feline protein; TRFP; leader A; leader B;
 KW epitope.
 OS Felis.
 FH Key Location/Qualifiers
 FT Peptide 1..22
 FT /label= leader_peptide
 PN WD9308280-A.
 PD 29-APR-1993.
 PF 16-OCT-1992; U08694.
 PR 16-OCT-1991; US-777859.
 PR 13-DEC-1991; US-807529.
 PA (IMMU-) IMMULOGIC PHARM CORP.
 PI Bond JF, Garman RD, Kuo M, Morgenstern JP, Morville M;
 PI Rogers BL;
 DR WPI; 93-152473/18.
 DR N-PSDB; Q41556.
 PT Recombitope peptide having T-cell stimulating activity - for the
 PT diagnosis and treatment of sensitivity to protein allergens,
 PT auto:antigens and protein antigens
 PS Disclosure; Fig 1; 73pp; English.
 CC Chains 1 and 2 of the TRFP have been recombinantly expressed in E.
 CC coli and purified. T cell epitope studies using overlapping peptide
 CC regions derived from the TRFP amino acids sequence were used to
 CC identify multiple T cell epitopes in each chain of TRFP.
 SQ Sequence 92 AA;
 SQ 9 A; 3 R; 4 N; 6 D; 0 B; 4 C; 2 Q; 7 E; 0 Z; 4 G; 0 H;
 SQ 4 I; 15 L; 7 K; 2 M; 1 F; 4 P; 2 S; 4 T; 2 W; 3 Y; 9 V;

Initial Score = 27 Optimized Score = 27 Significance = 15.21
 Residue Identity = 100% Matches = 27 Mismatches = 0

```

X      10      20
KALPVLLENARILKNCVDAKMT
|||||
MKGARVLVLLWAALLLIWGGNCEICPAVKRDVDFLTGTPDEYVEQVAQYKALPVLLENARILKNCVDAKMT
10      20      30      40      50      60      70

```

```

X
EEDKE
|||||
EEDKENALSLLDKIYTSPLC
X 80      90

```

9. US-08-300-510-2 (1-27)

R41983 Human T cell reactive feline protein A chain 1.

ID R41983 standard; Protein; 92 AA.
AC R41983;
DT 21-APR-1994 (first entry)
DE Human T cell reactive feline protein A chain 1.
KW Human; T cell; reactive; feline; protein; immune response; antigen;
KW tolerance; mammal; Dermatophagoides; Felis; Ambrosia; Lolium; Canis;
KW Cryptomeria; Alternaria; Alder; Betula; Quercus; Olea; Artemesia;
KW Plantago; Parietaria; Blattella; Apis; Periplaneta; autoantigen.
QS Homo sapiens.
FH Key Location/Qualifiers
FT Peptide 1..22
FT /note= "Signal peptide"
FT Protein 23..92
FT /note= "Mature protein"
PN W09319178-A.
PD 30-SEP-1993.
PF 25-MAR-1993; U02462.
PR 25-MAR-1992; US-857311.
PR 15-MAY-1992; US-884718.
PR 15-JAN-1993; US-006116.
PA (IMMU-) IMMUNOLOGIC PHARM CORP.
PI Briner TJ, Garman RD, Gefter ML, Greenstein JL, Kuo M;
PI Morville M;
DR WPI; 93-320744/40.
DR N-PSDB; 049533.
PT New peptide(s) for inducing tolerance - comprise one or more
PT epitope(s) of an allergen administered subcutaneously, for
PT treating sensitivity to cats, bees, etc.
PS Disclosure; Fig 1; 107pp; English.
CC The sequences given in R41983-84 represent chain 1 of human T cell
CC reactive feline proteins (TRFP) A and B respectively. Peptides
CC derived from TRFP may be used in a therapeutic composition which is
CC useful in treating diseases which involve an immune response to a
CC protein antigen. This composition may be used to induce tolerance
CC in a mammal to Dermatophagoides, Felis, Ambrosia, Lolium, Cryptomeria,
CC Alternaria, Alder, Betula, Quercus, Olea, Artemesia, Plantago,
CC Parietaria, Canis, Blattella, Apis, Periplaneta and to autoantigens
CC in humans.
SQ Sequence 92 AA;
SQ 9 A; 3 R; 4 N; 6 D; 0 B; 4 C; 2 Q; 7 E; 0 Z; 4 G; 0 H;
SQ 4 I; 15 L; 7 K; 2 M; 1 F; 4 P; 2 S; 4 T; 2 W; 3 Y; 9 V;

Initial Score = 27 Optimized Score = 27 Significance = 15.21
Residue Identity = 100% Matches = 27 Mismatches = 0
Gaps = 0 Conservative Substitutions = 0

```

X      10      20
KALPVLLENARILKNCVDAKMT

```

MKGARVLVLLWAALLLIWGGNCEICPAVKRDVDLFLTGTPDEYVEQVAQYKALPVLLENARILKNCVDAKMT
 10 20 30 40 50 60 70

X
 EEDKE
 |||||
 EEDKENALSLLDKIYTSPLC
 X 80 90

10. US-08-300-510-2 (1-27)
 R36540 TRFP chain 1 (with Leader B).

ID R36540 standard; Protein; 88 AA.
 AC R36540;
 DT 12-AUG-1993 (first entry)
 DE TRFP chain 1 (with Leader B).
 KW Human T cell reactive feline protein; TRFP; leader A; leader B;
 KW epitope.
 OS Felis.
 FH Key Location/Qualifiers
 FT Peptide 1..18
 FT /label= leader_peptide
 PN W09308280-A.
 PD 29-APR-1993.
 PF 16-OCT-1992; U08694.
 PR 16-OCT-1991; US-777859.
 PR 13-DEC-1991; US-807529.
 PA (IMMU-) IMMULOGIC PHARM CORP.
 PI Bond JF, Garman RD, Kuo M, Morgenstern JP, Morville M;
 PI Rogers BL;
 DR WPI; 93-152473/18.
 DR N-PSDB; Q41557.
 PT Recombitope peptide having T-cell stimulating activity - for the
 PT diagnosis and treatment of sensitivity to protein allergens,
 PT auto:antigens and protein antigens
 PS Disclosure; Fig 1; 73pp; English.
 CC Chains 1 and 2 of the TRFP have been recombinantly expressed in E.
 CC coli and purified. T cell epitope studies using overlapping peptide
 CC regions derived from the TRFP amino acids sequence were used to
 CC identify multiple T cell epitopes in each chain of TRFP.
 SQ Sequence 88 AA;
 SQ 11 A; 2 R; 3 N; 8 D; 0 B; 5 C; 2 Q; 7 E; 0 Z; 1 G; 0 H;
 SQ 3 I; 11 L; 6 K; 2 M; 1 F; 7 P; 2 S; 6 T; 0 W; 3 Y; 8 V;

Initial Score = 27 Optimized Score = 27 Significance = 15.21
 Residue Identity = 100% Matches = 27 Mismatches = 0
 Gaps = 0 Conservative Substitutions = 0

X 10 20
 KALPVLLENARILKNCVDAKMTTEEDK
 |||||
 MLDAALPPCPTVAATADCEICPAVKRDVDLFLTGTPDEYVEQVAQYKALPVLLENARILKNCVDAKMTTEEDK
 10 20 30 40 X 50 60 70

X
 E
 |
 ENALSLLDKIYTSPLC
 X 80

11. US-08-300-510-2 (1-27)
 R41984 Human T cell reactive feline protein B chain 1.

ID R41984 standard; Protein; 88 aa.
 AC R41984;
 DT 21-APR-1994 (first entry)
 DE Human T cell reactive feline protein B chain 1.
 KW Human; T cell; reactive; feline; protein; immune response; antigen;
 KW tolerance; mammal; Dermatophagoides; Felis; Ambrosia; Lolium; Canis;
 KW Cryptomeria; Alternaria; Alder; Betula; Quercus; Olea; Artemesia;
 KW Plantago; Parietaria; Blattella; Apis; Periplaneta; autoantigen.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT Peptide 1..17
 FT /note= "Signal peptide"
 FT Protein 18..88
 FT /note= "Mature protein"
 PN WD9319178-A.
 PD 30-SEP-1993.
 PF 25-MAR-1993; U02462.
 PR 25-MAR-1992; US-857311.
 PR 15-MAY-1992; US-884718.
 PR 15-JAN-1993; US-006116.
 PA (IMMU-) IMMUNOLOGIC PHARM CORP.
 PI Briner TJ, Garman RD, Geftler ML, Greenstein JL, Kuo M;
 PI Morville M;
 DR WPI; 93-320744/40.
 DR N-PSDB; Q49534.
 PT New peptide(s) for inducing tolerance - comprise one or more
 PT epitope(s) of an allergen administered subcutaneously, for
 FT treating sensitivity to cats, bees, etc.
 PS Disclosure; Fig 1; 107pp; English.
 CC The sequences given in R41983-84 represent chain 1 of human T cell
 CC reactive feline proteins (TRFP) A and B respectively. Peptides
 CC derived from TRFP may be used in a therapeutic composition which is
 CC useful in treating diseases which involve an immune response to a
 CC protein antigen. This composition may be used to induce tolerance
 CC in a mammal to Dermatophagoides, Felis, Ambrosia, Lolium, Cryptomeria,
 CC Alternaria, Alder, Betula, Quercus, Olea, Artemesia, Plantago,
 CC Parietaria, Canis, Blattella, Apis, Periplaneta and to autoantigens
 CC in humans.
 SQ Sequence 88 AA;
 SQ 11 A; 2 R; 3 N; 8 D; 0 B; 5 C; 2 Q; 7 E; 0 Z; 1 G; 0 H;
 SQ 3 I; 11 L; 6 K; 2 M; 1 F; 7 P; 2 S; 6 T; 0 W; 3 Y; 8 V;

Initial Score = 27 Optimized Score = 27 Significance = 15.21
 Residue Identity = 100% Matches = 27 Mismatches = 0
 Gaps = 0 Conservative Substitutions = 0

X 10 20
 KALPVVLENARILKNCVDAKMTEEDK
 |||||
 MLDAALPPCPTVAATADCEICPAVKRDVDFLTGTPDEYVEQVAGYKALPVVLENARILKNCVDAKMTEEDK
 10 20 30 40 X 50 60 70

X
 E
 I
 ENALSLLDKIYTSPLC
 X 80

12. US-08-300-510-2 (1-27)
 R13139 B.burgdorferi strain PKo p100 gene.

ID R13139 standard; Protein; 663 AA.
 AC R13139;
 DT 27-SEP-1991 (first entry)
 DE B.burgdorferi strain PKo p100 gene.

KW Lyme borreliosis; vaccine; flagellin; ss.
 OS Borrelia burgdorferi.
 PN W09109870-A.
 PD 11-JUL-1991.
 PF 21-DEC-1990; E02282.
 PR 22-DEC-1989; DE-942728.
 PR 13-JUN-1990; DE-018988.
 PA (MIKR-) MIKROGEN MOLEKULARB.
 PI Fuchs R, Wilske B, Preac-Mursic V, Motz M, Soutschek E.
 DR WPI; 91-222844/30.
 PT New Borrelia burgdorferi proteins - useful as immunoassay
 PT reagents and antigens for vaccine prodn.
 PS Claim 11; Page 49; 68pp; German.
 CC Protein p100 was isolated from a B.burgdorferi cell lysate and the N-
 CC terminal amino acid sequence was determined. A probe pool was
 CC synthesised and used to screen a B.burgdorferi cDNA library. A clone
 CC contg. the 5' 346 nucleotides of the p100 coding sequence was
 CC identified and sequenced. Cloning the entire gene allowed the p100
 CC amino acid sequence to be deduced.
 CC See also Q12744-Q12747, Q13297-8 and R13140-R13142.
 SQ Sequence 663 AA;
 SQ 28 A; 19 R; 46 N; 64 D; 0 B; 0 C; 29 Q; 65 E; 0 Z; 21 G; 4 H;
 SQ 52 I; 70 L; 79 K; 4 M; 25 F; 16 P; 59 S; 22 T; 2 W; 19 Y; 39 V;

Initial Score = 9 Optimized Score = 9 Significance = 4.26
 Residue Identity = 33% Matches = 9 Mismatches = 18
 Gaps = 0 Conservative Substitutions = 0

IDLDKAQQKLDFAEDNLDIQRTVREKLQENINETNKEKNLPKPGDVSSPKVDKQLQIKESLEDLQEQLKEA
 300 310 320 330 340 350 360 370

 X 10 20 X
 KALPVLLENARILKNCVDAKMTEEDKE
 ||| | ||||
 SDENQKREIEKQIEIKKNDEELFKNKDHLKALDLKQELNSKASSKEKIEGEEEDKELDSKKNLEPVSEADKVD
 380 390 400 410 420 X 430 440

 KISKSNNEVSKLSPLDEPSYSDIDSKEGVDNKDVLQKTKPQVESQPTSLNEDLIDVSDSSNPVFLEVID
 450 460 470 480 490 500 510

 PITNLGTLQLI
 520

13. US-08-300-510-2 (1-27)
 R44489 Sequence of all or part of a mammalian calmodulin-

ID R44489 standard; Protein; 1429 AA.
 AC R44489;
 DT 19-JUN-1994 (first entry)
 DE Sequence of all or part of a mammalian calmodulin-dependent
 DE nitric oxide synthase (NOS).
 KW Calmodulin-dependent nitric oxide synthase; NOS;
 KW immunohistochemical reagent; antibody; assay.
 OS Homo sapiens.
 PN US5268465-A.
 PD 07-DEC-1993.
 PF 18-JAN-1991; 642002.
 PR 18-JAN-1991; US-642002.
 PA (UYJD) UNIV JOHNS HOPKINS.
 PI Brett DS, Snyder SH;
 DR WPI; 93-404061/50.
 DR N-PSDB; Q53403.
 PT DNA encoding mammalian, calmodulin-dependent nitric oxide
 PT synthase - used to raise antibodies which localise NOS in the
 PT body, useful as immuno:histochemical reagent

PS Disclosure; columns 7-16; 20pp; English.
 CC Degenerate oligonucleotide (OG) primers of 21 nucleotides were
 CC constructed, based on the seven AAs at the carboxyl and amino
 CC termini of each of the two longest trypsin peptides of NOS enzyme
 CC (18 and 17 AAs). These OGs were used in a PCR reaction to construct
 CC two non-degenerate OG primers. The two non-degenerate primers were
 CC used in a further PCR reaction to obtain a larger polynucleotide
 CC probe. A 600 bp amplified prod. was obtd. and random prime-labelled
 CC with (32) P-ATP to screen a commercially obtd. rat brain cDNA
 CC library. Eight overlapping independent clones were isolated and
 CC sequenced by ds dideoxy sequencing. A 4 kb ORF encoding a 150 kD
 CC protein (corresp. to the mol. wt. on NOS) was revealed. A flavin-
 CC binding consensus sequence was observed in the AA sequence.
 SQ Sequence 1429 AA:
 SQ 87 A; 80 R; 61 N; 82 D; 0 B; 24 C; 68 Q; 95 E; 0 Z; 101 G; 41 H;
 SQ 68 I; 124 L; 90 K; 29 M; 59 F; 81 P; 100 S; 78 T; 21 W; 41 Y; 99 V;

Initial Score = 9 Optimized Score = 9 Significance = 4.26
 Residue Identity = 33% Matches = 9 Mismatches = 18
 Gaps = 0 Conservative Substitutions = 0

GDDVNIIEKPNNLSISNDRSWKRNFRLTYVAEAPDLTQGLSNVHKKRVSAARLLSRQNLQSPKFSRSTIFVR
 950 960 970 980 990 1000 1010
 X 10 20 X
 KALPVVLENARILKNCVDAKMTEEDKE
 || || | | ||
 LHTNGNQELQYQPGDHLGVFPGNHEDLVNALIERLEDAPPANHVVVKVEMLEERNTALGVISNWKDESRLPPC
 1020 1030 1040 X 1050 1060 1070 1080
 TIFQAFKYYLDITTPPTPLQLQGFASLATNEKEKQRLVLKGLQEEYEEWKWGKNPTMVEVLEEFPSIQMPA
 1090 1100 1110 1120 1130 1140 1150 1160
 TLLLTQLSLLQ
 1170

14. US-08-300-510-2 (1-27)
 R15140 Vascular injury affinity peptide.

ID R15140 standard; Protein; 18 AA.
 AC R15140;
 DT 18-FEB-1992 (first entry)
 DE Vascular injury affinity peptide.
 KW Low density lipoprotein; atherosclerosis.
 OS Synthetic.
 PN WD9116919-A.
 PD 14-NOV-1991.
 PF 02-MAY-1991; U03026.
 PR 03-MAY-1990; US-518215.
 PR 03-MAY-1990; US-518142.
 PA (NEW-) NEW ENGLAND DEACON.
 PI Lees RS, Lees AM, Fischman A, Shih IL, Findeis MA;
 DR WPI; 91-353525/48.
 PT Synthetic peptide(s) comprising amphiphilic domain of apoA-I -
 PT used to diagnose vascular injury or disease or inhibit binding of
 PT low density lipoprotein to vascular walls in treating
 PT atherosclerosis
 PS Disclosure; Page 8; 66pp; English.
 CC The amino acid sequence is that of a synthetic peptide (opt. labelled)
 CC which is used to detect injuries in the vascular system, esp. athero-
 CC sclerosis in its early stages before it causes stenosis and blood
 CC flow disturbances. It can also be used to inhibit binding of low
 CC density lipoprotein (LDL) to vascular walls, i.e. to prevent or
 CC alleviate atherosclerosis. It is easy to prepare on a large scale
 CC and allows vascular regions to be located non-invasively without

CC complex equipment or highly skilled personnel. See also R15126-R15137.

SQ Sequence 18 AA;
SQ 6 A; 1 R; 1 N; 0 D; 0 B; 0 C; 0 Q; 2 E; 0 Z; 1 G; 0 H;
SQ 0 I; 4 L; 2 K; 0 M; 0 F; 0 P; 0 S; 0 T; 0 W; 1 Y; 0 V;

Initial Score = 8 Optimized Score = 8 Significance = 3.65
Residue Identity = 44% Matches = 8 Mismatches = 10
Gaps = 0 Conservative Substitutions = 0

```

X      10      20
KALPVLLENARILKNCVDAKMTEEDKE
      || || || || ||
YKLALAAARLLANAEGAK
X      10      X
```

15. US-08-300-510-2 (1-27)
R15127 Vascular injury affinity peptide.

ID R15127 standard; Protein; 18 AA.
AC R15127;
DT 18-FEB-1992 (first entry)
DE Vascular injury affinity peptide.
KW Low density lipoprotein; atherosclerosis.
OS Synthetic.
PN W09116919-A.
PD 14-NOV-1991.
PF 02-MAY-1991; U03026.
PR 03-MAY-1990; US-518215.
PR 03-MAY-1990; US-518142.
PA (NEW-) NEW ENGLAND DEACON.
PI Lees RS, Lees AM, Fischman A, Shih IL, Findeis MA;
DR WPI; 91-353525/48.
PT Synthetic peptide(s) comprising amphiphilic domain of apoA-I -
PT used to diagnose vascular injury or disease or inhibit binding of
PT low density lipoprotein to vascular walls in treating
PT atherosclerosis
PS Claim 1; Page 44; 66pp; English.
CC The amino acid sequence is that of a synthetic peptide (opt. labelled)
CC which is used to detect injuries in the vascular system, esp. athero-
CC sclerosis in its early stages before it causes stenosis and blood
CC flow disturbances. It can also be used to inhibit binding of low
CC density lipoprotein (LDL) to vascular walls, i.e. to prevent or
CC alleviate atherosclerosis. It is easy to prepare on a large scale
CC and allows vascular regions to be located non-invasively without
CC complex equipment or highly skilled personnel. See also R15126-R15140.
SQ Sequence 18 AA;
SQ 6 A; 1 R; 1 N; 0 D; 0 B; 0 C; 0 Q; 2 E; 0 Z; 1 G; 0 H;
SQ 0 I; 4 L; 2 K; 0 M; 0 F; 0 P; 0 S; 0 T; 0 W; 1 Y; 0 V;

Initial Score = 8 Optimized Score = 8 Significance = 3.65
Residue Identity = 44% Matches = 8 Mismatches = 10
Gaps = 0 Conservative Substitutions = 0

```

X      10      20
KALPVLLENARILKNCVDAKMTEEDKE
      || || || || ||
YKLALAAARLLANAEGAK
X      10      X
```

=> fil reg

FILE 'REGISTRY' ENTERED AT 11:20:30 ON 24 MAR 95
USE IS SUBJECT TO THE TERMS OF YOUR CUSTOMER AGREEMENT
COPYRIGHT (C) 1995 American Chemical Society (ACS)

STRUCTURE FILE UPDATES: 17 MAR 95 HIGHEST RN 161594-17-8
DICTIONARY FILE UPDATES: 23 MAR 95 HIGHEST RN 161594-17-8

TSCA INFORMATION NOW CURRENT THROUGH MAY 1994

Please note that search-term pricing does apply when
conducting SmartSELECT searches.

=> d que 12

L2 13 SEA FILE=REGISTRY KRDVDLFLTGTGPDEYVEQVAQYKALPV|KALPVLLENAR
ILKNCVDAKMTEEDKE/SQSP 8Q 10 1 + 2

=> d ide can 12 1-13

L2 ANSWER 1 OF 13 REGISTRY COPYRIGHT 1995 ACS
RN 152416-30-3 REGISTRY
CN Allergen Fel d I (Felis catus) (9CI) (CA INDEX NAME)
FS PROTEIN SEQUENCE
MF Unspecified
CI MAN
SR CA

*** STRUCTURE DIAGRAM IS NOT AVAILABLE ***
*** USE 'SQD' OR 'SQIDE' FORMATS TO DISPLAY SEQUENCE ***
1 REFERENCES IN FILE CA (1967 TO DATE)

REFERENCE 1: 120:75353

L2 ANSWER 2 OF 13 REGISTRY COPYRIGHT 1995 ACS
RN 149119-99-3 REGISTRY
CN 47-73-Glycoprotein TRFP (Felis catus chain 1 isoform B protein moiety reduced), N-(L-methionylglycyl-L-histidyl-L-histidyl-L-histidyl-L-histidyl-L-histidyl-L-.alpha.-glutamyl-L-phenylalanyl-L-leucyl-L-valyl-L-prolyl-L-arginylglycyl-L-seryl)-, (73.fwdarw.14')-protein with 14-39-allergen Fel dI (Felis catus chain 2 protein moiety reduced) (39'.fwdarw.25')-protein with 25-51-glycoprotein TRFP (Felis catus chain 1 isoform B protein moiety reduced) (9CI) (CA INDEX NAME)
FS PROTEIN SEQUENCE
MF Unspecified
CI MAN
SR CA

*** STRUCTURE DIAGRAM IS NOT AVAILABLE ***
*** USE 'SQD' OR 'SQIDE' FORMATS TO DISPLAY SEQUENCE ***
1 REFERENCES IN FILE CA (1967 TO DATE)

REFERENCE 1: P 119:93527

L2 ANSWER 3 OF 13 REGISTRY COPYRIGHT 1995 ACS
RN 149119-95-9 REGISTRY
CN 69-87-Glycoprotein TRFP (Felis catus chain 1 isoform B protein moiety reduced), (87.fwdarw.47')-protein with 47-73-glycoprotein TRFP (Felis catus chain 1 isoform B protein moiety reduced) (73'.fwdarw.14'')-protein with 14-39-allergen Fel dI (Felis catus chain 2 protein moiety reduced) (39''.fwdarw.25'')-protein with 25-51-glycoprotein TRFP (Felis catus chain 1 isoform B protein moiety reduced) (51'''.fwdarw.74''''')-protein with 74-92-allergen Fel dI (Felis catus chain 2 protein moiety reduced) (9CI) (CA INDEX NAME)
FS PROTEIN SEQUENCE
MF Unspecified
CI MAN
SR CA

*** STRUCTURE DIAGRAM IS NOT AVAILABLE ***
*** USE 'SQD' OR 'SQIDE' FORMATS TO DISPLAY SEQUENCE ***
1 REFERENCES IN FILE CA (1967 TO DATE)

REFERENCE 1: P 119:93527

L2 ANSWER 4 OF 13 REGISTRY COPYRIGHT 1995 ACS
RN 149119-94-8 REGISTRY
CN 47-73-Glycoprotein TRFP (Felis catus chain 1 isoform B protein moiety reduced), (73.fwdarw.14')-protein with 14-39-allergen Fel dI (Felis catus chain 2 protein moiety reduced) (39'.fwdarw.25'')-protein with 25-51-glycoprotein TRFP (Felis catus chain 1 isoform B protein moiety reduced) (9CI) (CA INDEX NAME)
FS PROTEIN SEQUENCE
MF Unspecified
CI MAN
SR CA

*** STRUCTURE DIAGRAM IS NOT AVAILABLE ***
*** USE 'SQD' OR 'SQIDE' FORMATS TO DISPLAY SEQUENCE ***
1 REFERENCES IN FILE CA (1967 TO DATE)

REFERENCE 1: P 119:93527

L2 ANSWER 5 OF 13 REGISTRY COPYRIGHT 1995 ACS
RN 136796-97-9 REGISTRY
CN Glycoprotein TRFP (Felis catus chain 1 isoform B protein moiety reduced) (9CI) (CA INDEX NAME)
OTHER NAMES:
CN Allergen Fel dI (Felis catus chain 1 isoform B precursor protein moiety reduced)
CN Leader B-human T cell-reactive feline protein chain 1 (cat)

FS PROTEIN SEQUENCE
MF C422 H685 N105 O135 S7
CI MAN
SR CA
LC STN Files: CA
DES 5:ALL,L

*** STRUCTURE DIAGRAM IS NOT AVAILABLE ***
*** USE 'SQD' OR 'SQIDE' FORMATS TO DISPLAY SEQUENCE ***
4 REFERENCES IN FILE CA (1967 TO DATE)

REFERENCE 1: P 120:75451
REFERENCE 2: P 119:93527
REFERENCE 3: 118:122542
REFERENCE 4: P 115:205920

L2 ANSWER 6 OF 13 REGISTRY COPYRIGHT 1995 ACS
RN 136796-96-8 REGISTRY
CN Glycoprotein TRFP (Felis catus chain 1 isoform A protein moiety reduced) (9CI) (CA INDEX NAME)

OTHER NAMES:

CN Allergen Fel dI (Felis catus chain 1 isoform A precursor protein moiety reduced)
CN Leader A-human T cell-reactive feline protein chain 1 (cat)
FS PROTEIN SEQUENCE
MF C461 H748 N116 O134 S6
CI MAN
SR CA
LC STN Files: CA
DES 5:ALL,L

*** STRUCTURE DIAGRAM IS NOT AVAILABLE ***
*** USE 'SQD' OR 'SQIDE' FORMATS TO DISPLAY SEQUENCE ***
4 REFERENCES IN FILE CA (1967 TO DATE)

REFERENCE 1: P 120:75451
REFERENCE 2: P 119:93527
REFERENCE 3: 118:122542
REFERENCE 4: P 115:205920

L2 ANSWER 7 OF 13 REGISTRY COPYRIGHT 1995 ACS
RN 136796-93-5 REGISTRY
CN 23-92-Glycoprotein TRFP (Felis catus chain 1 isoform A protein moiety reduced) (9CI) (CA INDEX NAME)

OTHER NAMES:

CN Allergen Fel dI (Felis catus chain 1 protein moiety reduced)
FS PROTEIN SEQUENCE
MF C348 H565 N87 O111 S4
CI MAN
SR CA
LC STN Files: CA
DES 5:ALL,L

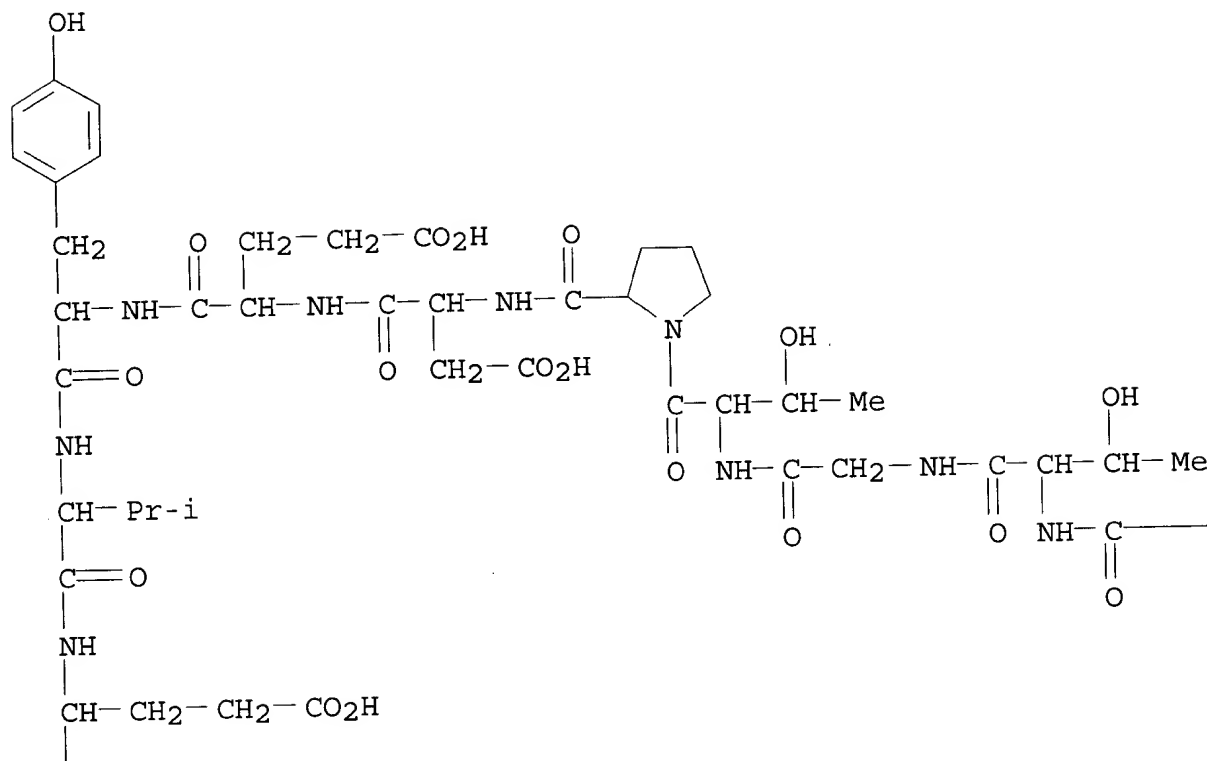
*** STRUCTURE DIAGRAM IS NOT AVAILABLE ***
*** USE 'SQD' OR 'SQIDE' FORMATS TO DISPLAY SEQUENCE ***
2 REFERENCES IN FILE CA (1967 TO DATE)

REFERENCE 1: 118:122542

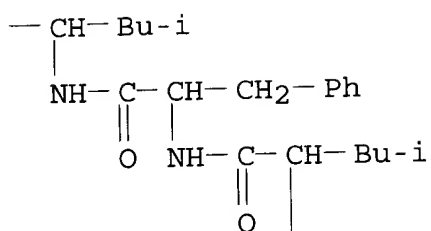
REFERENCE 2: P 115:205920

L2 ANSWER 8 OF 13 REGISTRY COPYRIGHT 1995 ACS
RN 136414-74-9 REGISTRY
CN L-Valine, L-alanyl-L-valyl-L-lysyl-L-arginyl-L-.alpha.-aspartyl-L-
valyl-L-.alpha.-aspartyl-L-leucyl-L-phenylalanyl-L-leucyl-L-
threonylglycyl-L-threonyl-L-prolyl-L-.alpha.-aspartyl-L-.alpha.-
glutamyl-L-tyrosyl-L-valyl-L-.alpha.-glutamyl-L-glutaminyl-L-valyl-L-
alanyl-L-glutaminyl-L-tyrosyl-L-lysyl-L-alanyl-L-leucyl-L-prolyl-
(9CI) (CA INDEX NAME)
FS PROTEIN SEQUENCE
MF C149 H234 N36 O46
SR CA
LC STN Files: CA
DES 5:ALL,L

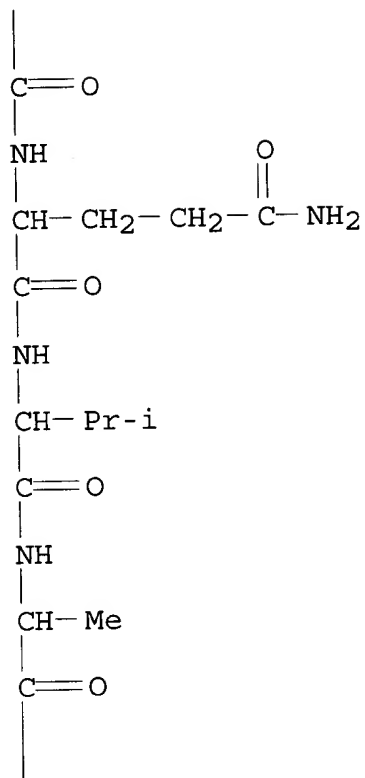
PAGE 1-A



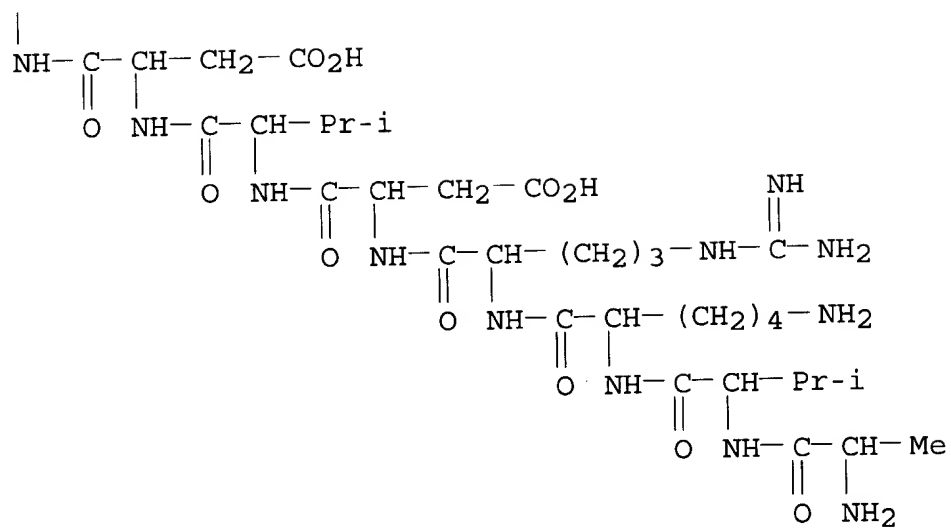
PAGE 1-B



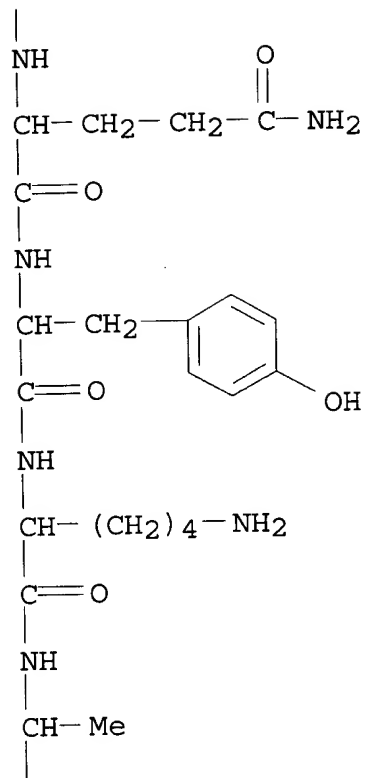
PAGE 2-A



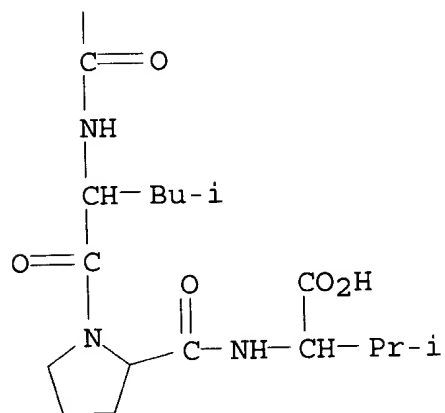
PAGE 2-B



PAGE 3-A



PAGE 4-A

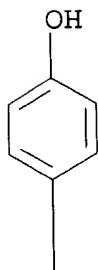


1 REFERENCES IN FILE CA (1967 TO DATE)

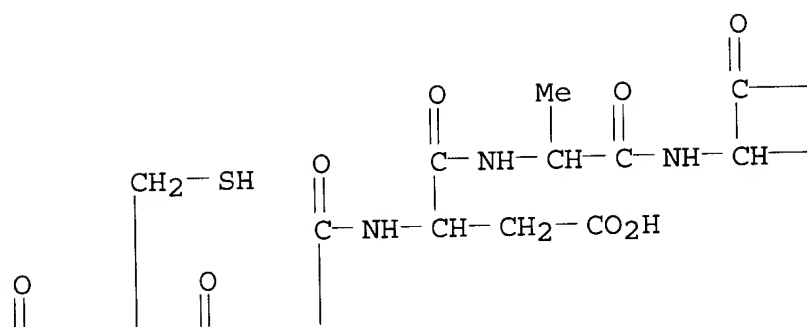
REFERENCE 1: P 115:205920

L2 ANSWER 9 OF 13 REGISTRY COPYRIGHT 1995 ACS
RN 136380-73-9 REGISTRY
CN L-Glutamic acid, L-tyrosyl-L-lysyl-L-alanyl-L-leucyl-L-prolyl-L-
valyl-L-valyl-L-leucyl-L-.alpha.-glutamyl-L-asparaginyl-L-alanyl-L-
arginyl-L-isoleucyl-L-leucyl-L-lysyl-L-asparaginyl-L-cysteinyl-L-
valyl-L-.alpha.-aspartyl-L-alanyl-L-lysyl-L-methionyl-L-threonyl-L-
.alpha.-glutamyl-L-.alpha.-glutamyl-L-.alpha.-aspartyl-L-lysyl-
(9CI) (CA INDEX NAME)
FS PROTEIN SEQUENCE
MF C140 H235 N37 O45 S2
SR CA
LC STN Files: CA
DES 5:ALL,L

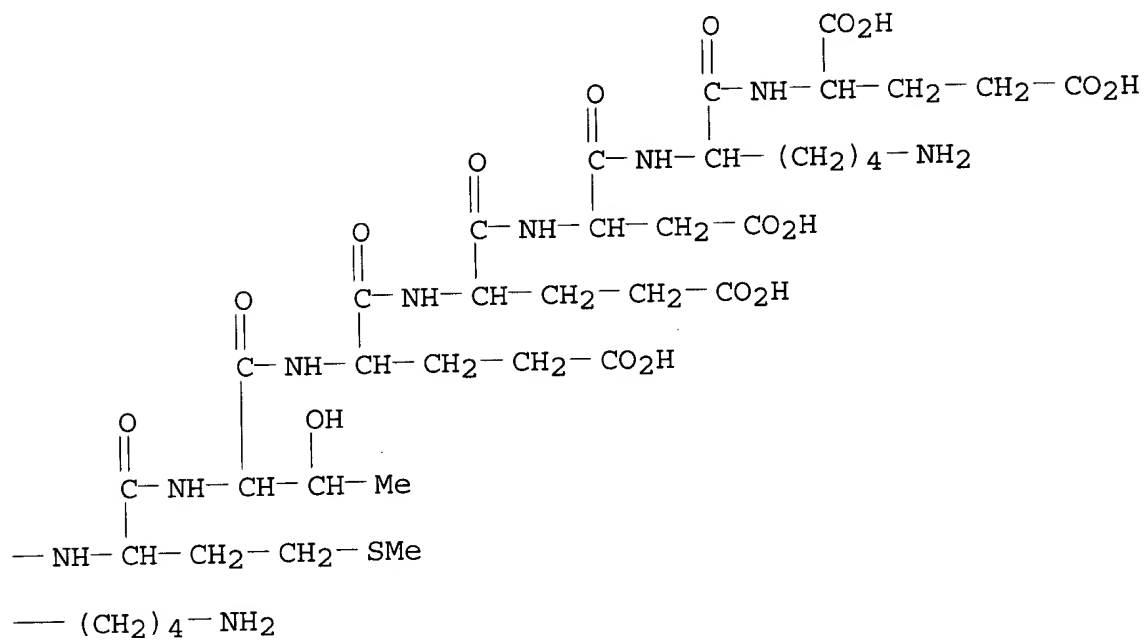
PAGE 1-A



PAGE 1-B

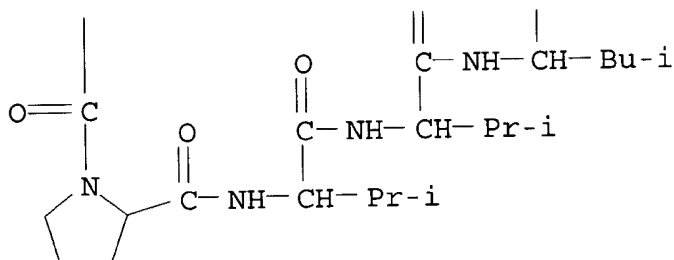


PAGE 1-C



$$\begin{array}{c}
 \text{CH}_2 \\
 | \\
 \text{CH}-\text{NH}_2 \\
 | \\
 \text{C}=\text{O} \\
 | \\
 \text{NH} \\
 | \\
 \text{CH}-(\text{CH}_2)_4-\text{NH}_2 \\
 | \\
 \text{C}=\text{O} \\
 | \\
 \text{NH} \\
 | \\
 \text{CH}-\text{Me} \\
 | \\
 \text{C}=\text{O} \\
 | \\
 \text{NH} \\
 | \\
 \text{CH}-\text{Bu-i}
 \end{array}$$
$$\begin{array}{c}
 \text{---} \text{CH} \text{---} \text{CH} \text{---} \text{Et} \\
 | \quad | \\
 \text{O} \quad \text{Me} \\
 || \quad | \\
 \text{C} \text{---} \text{NH} \text{---} \text{CH} \text{---} \text{Bu-i} \\
 | \\
 \text{O} \\
 || \\
 \text{C} \text{---} \text{NH} \text{---} \text{CH} \text{---} (\text{CH}_2)_4 \text{---} \text{NH}_2 \\
 | \\
 \text{O} \quad \text{O} \\
 || \quad || \\
 \text{C} \text{---} \text{NH} \text{---} \text{CH} \text{---} \text{CH}_2 \text{---} \text{C} \text{---} \text{NH}_2 \\
 | \quad | \quad | \\
 \text{O} \quad \text{O} \quad \text{O} \\
 || \quad || \quad || \\
 \text{C} \text{---} \text{NH} \text{---} \text{CH} \text{---} \text{C} \text{---} \text{NH} \text{---} \text{CH} \text{---} \text{Pr-i}
 \end{array}$$

PAGE 3-A

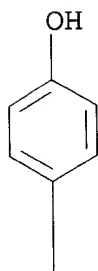


1 REFERENCES IN FILE CA (1967 TO DATE)

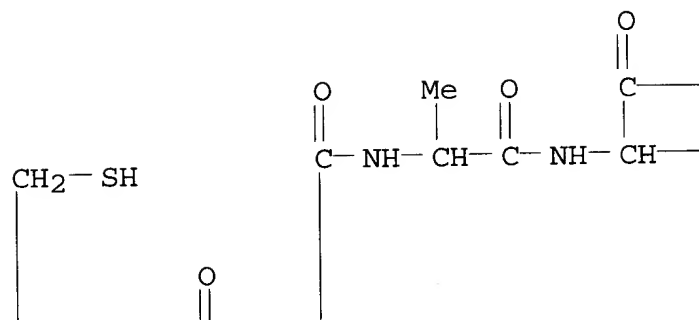
REFERENCE 1: P 115:205920

L2 ANSWER 10 OF 13 REGISTRY COPYRIGHT 1995 ACS
 RN 136380-72-8 REGISTRY
 CN L-Glutamic acid, L-alanyl-L-glutaminyl-L-tyrosyl-L-lysyl-L-alanyl-L-leucyl-L-prolyl-L-valyl-L-valyl-L-leucyl-L-.alpha.-glutamyl-L-asparaginyl-L-alanyl-L-arginyl-L-isoleucyl-L-leucyl-L-lysyl-L-asparaginyl-L-cysteinyl-L-valyl-L-.alpha.-aspartyl-L-alanyl-L-lysyl-L-methionyl-L-threonyl-L-.alpha.-glutamyl-L-.alpha.-glutamyl-L-.alpha.-aspartyl-L-lysyl- (9CI) (CA INDEX NAME)
 FS PROTEIN SEQUENCE
 MF C148 H248 N40 O48 S2
 SR CA
 LC STN Files: CA
 DES 5:ALL,L

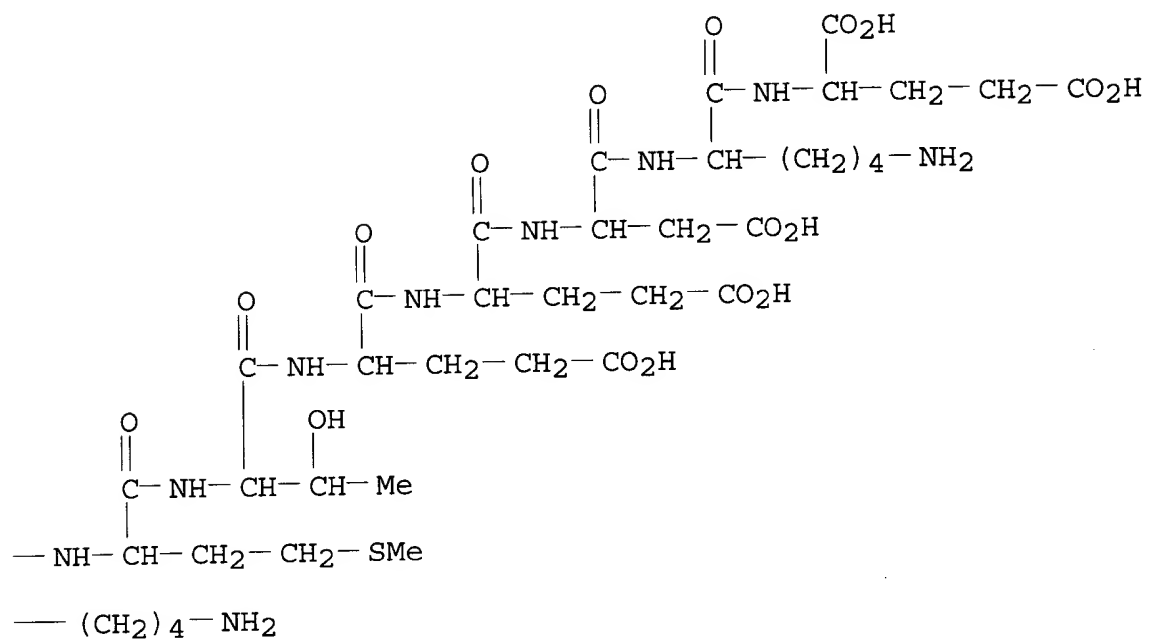
PAGE 1-A



PAGE 1-B



PAGE 1-C



$$\begin{array}{c}
 | \\
 \text{CH}_2 \\
 | \\
 \text{CH}-\text{NH}-\text{C}(=\text{O})-\text{CH}-\text{CH}_2-\text{CH}_2-\text{C}(=\text{O})-\text{NH}_2 \\
 | \quad | \quad | \quad | \quad | \\
 \text{NH}-\text{C}(=\text{O})-\text{CH}-\text{Me} \\
 | \\
 \text{NH}_2
 \end{array}$$

[illegible][illegible]

REFERENCE 1: P 115:205920

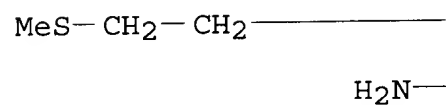
L2 ANSWER 11 OF 13 REGISTRY COPYRIGHT 1995 ACS
RN 136380-69-3 REGISTRY
CN L-Glutamic acid, L-lysyl-L-alanyl-L-leucyl-L-prolyl-L-valyl-L-valyl-L-leucyl-L-.alpha.-glutamyl-L-asparaginyl-L-alanyl-L-arginyl-L-
isoleucyl-L-leucyl-L-lysyl-L-asparaginyl-L-cysteinyl-L-valyl-L-
.alpha.-aspartyl-L-alanyl-L-lysyl-L-methionyl-L-threonyl-L-.alpha.-
glutamyl-L-.alpha.-glutamyl-L-.alpha.-aspartyl-L-lysyl- (9CI) (CA

INDEX NAME)

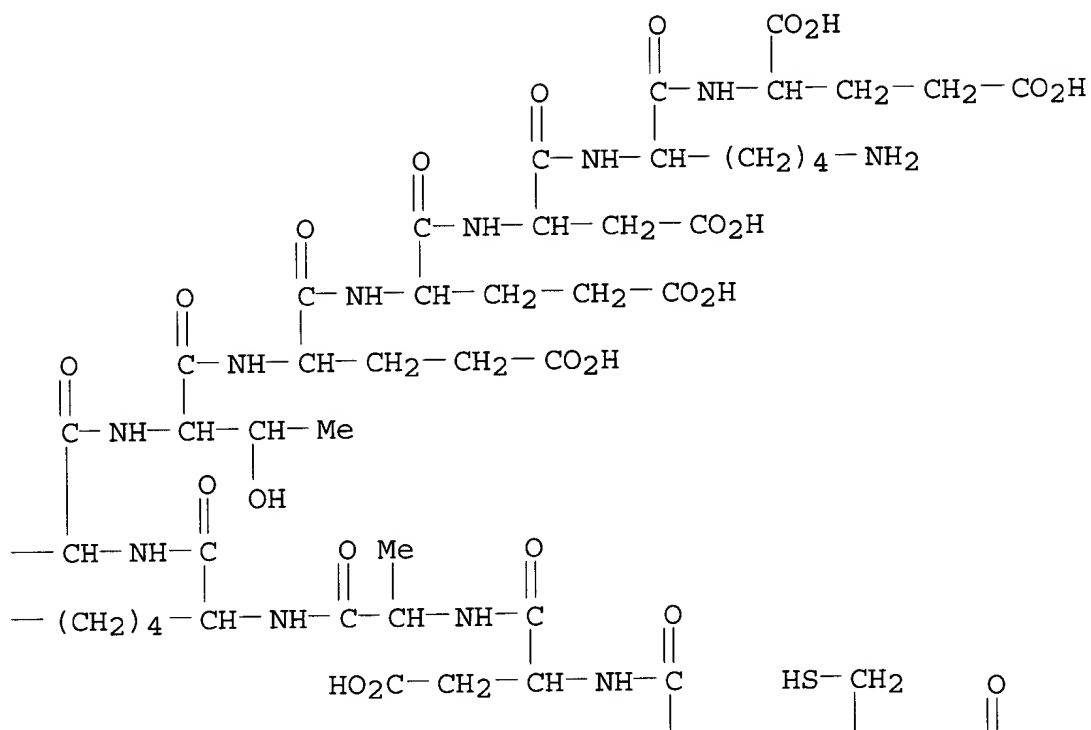
OTHER NAMES:

CN Human T cell-reactive feline protein chain 1 (29-55) (cat synthetic)
FS PROTEIN SEQUENCE
MF C131 H226 N36 O43 S2
SR CA
LC STN Files: CA
DES 5:ALL,L

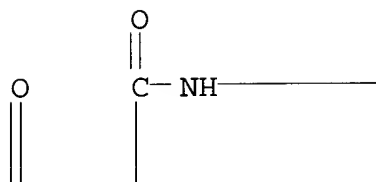
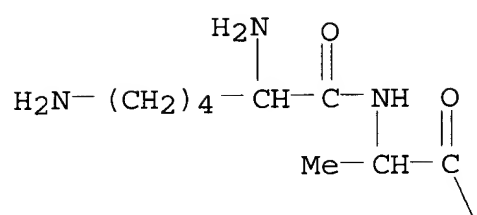
PAGE 1-A



PAGE 1-B



PAGE 2-A

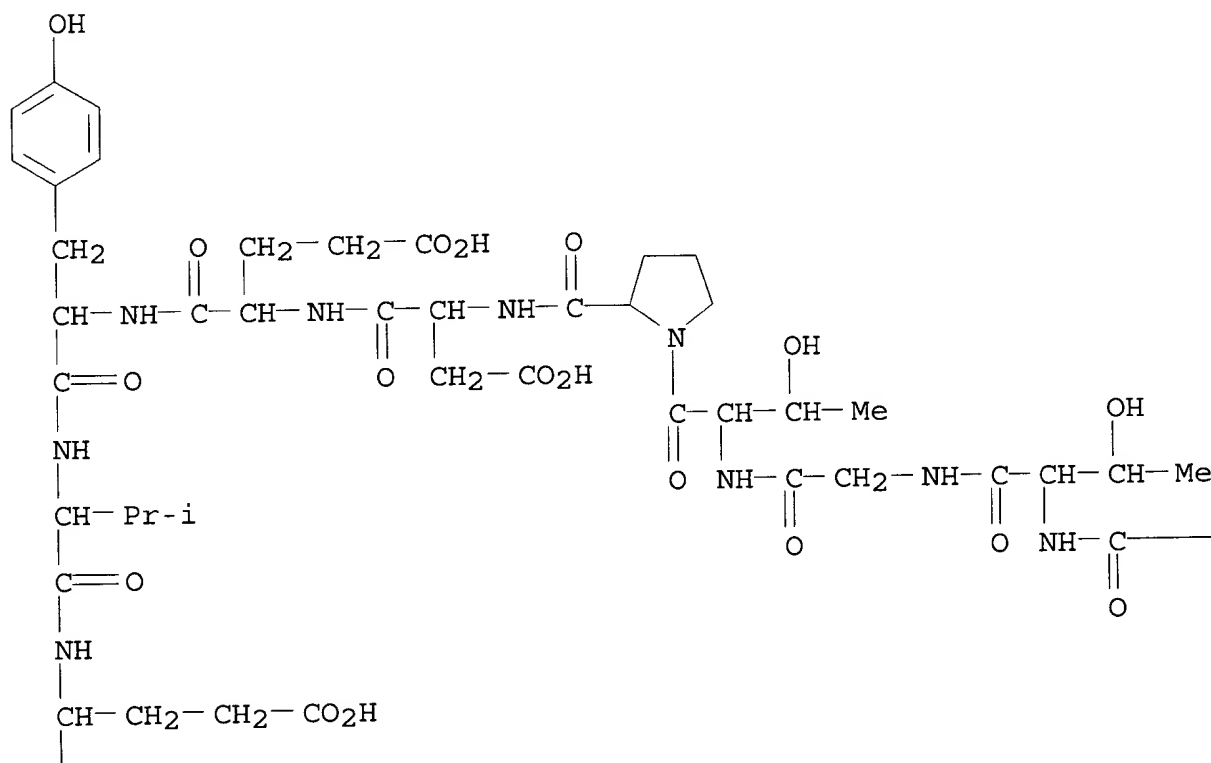


[illegible][illegible]

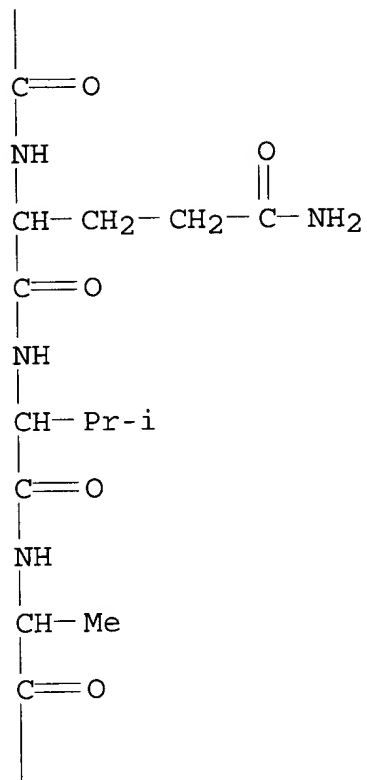
REFERENCE	1:		121:228217
REFERENCE	2:	P	120:75451
REFERENCE	3:	P	119:93527
REFERENCE	4:	P	115:205920

L2 ANSWER 12 OF 13 REGISTRY COPYRIGHT 1995 ACS
 RN 136380-56-8 REGISTRY
 CN L-Valine, L-lysyl-L-arginyl-L-.alpha.-aspartyl-L-valyl-L-.alpha.-
 aspartyl-L-leucyl-L-phenylalanyl-L-leucyl-L-threonylglycyl-L-
 threonyl-L-prolyl-L-.alpha.-aspartyl-L-.alpha.-glutamyl-L-tyrosyl-L-
 valyl-L-.alpha.-glutamyl-L-glutamyl-L-valyl-L-alanyl-L-glutamyl-L-
 L-tyrosyl-L-lysyl-L-alanyl-L-leucyl-L-prolyl- (9CI) (CA INDEX NAME)
 FS PROTEIN SEQUENCE
 MF C141 H220 N34 O44
 SR CA
 LC STN Files: CA
 DES 5:ALL,L

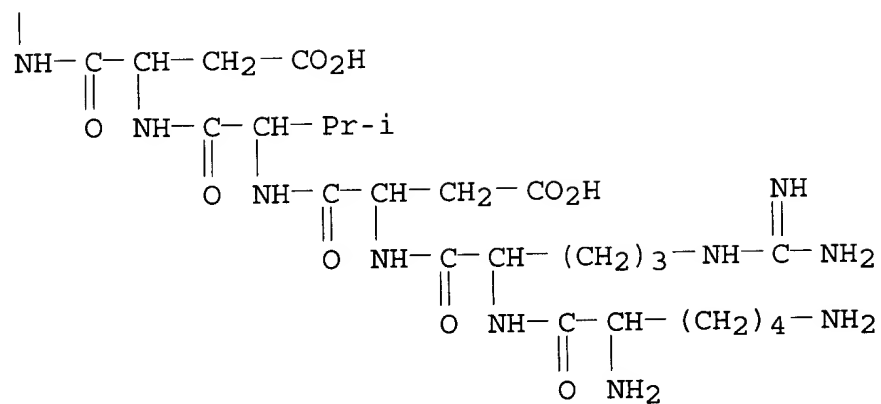
PAGE 1-A



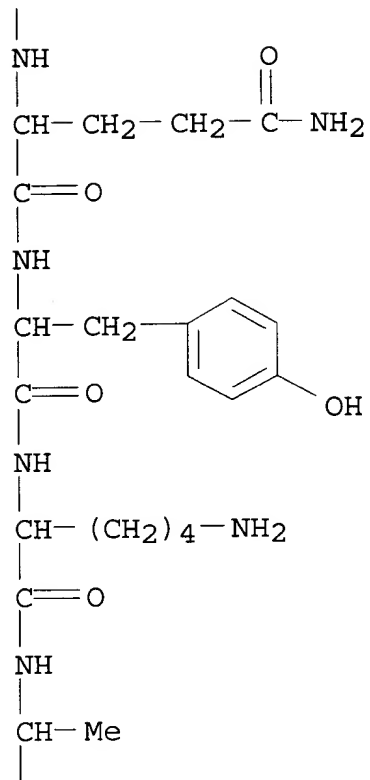
PAGE 2-A



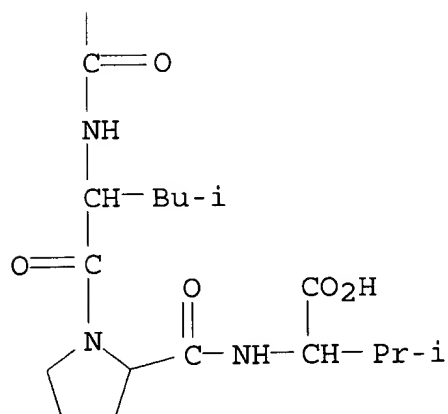
PAGE 2-B



PAGE 3-A



PAGE 4-A



3 REFERENCES IN FILE CA (1967 TO DATE)
 1 REFERENCES TO NON-SPECIFIC DERIVATIVES IN FILE CA

REFERENCE 1: 121:228217

REFERENCE 2: P 120:75451

REFERENCE 3: P 115:205920

L2 ANSWER 13 OF 13 REGISTRY COPYRIGHT 1995 ACS

RN 136380-55-7 REGISTRY

L-Valine, L-valyl-L-lysyl-L-arginyl-L-.alpha.-aspartyl-L-valyl-L-
 .alpha.-aspartyl-L-leucyl-L-phenylalanyl-L-leucyl-L-threonylglycyl-L-
 threonyl-L-prolyl-L-.alpha.-aspartyl-L-.alpha.-glutamyl-L-tyrosyl-L-
 valyl-L-.alpha.-glutamyl-L-glutaminyl-L-valyl-L-alanyl-L-glutaminyl-
 L-tyrosyl-L-lysyl-L-alanyl-L-leucyl-L-prolyl- (9CI) (CA INDEX NAME)

FS PROTEIN SEQUENCE

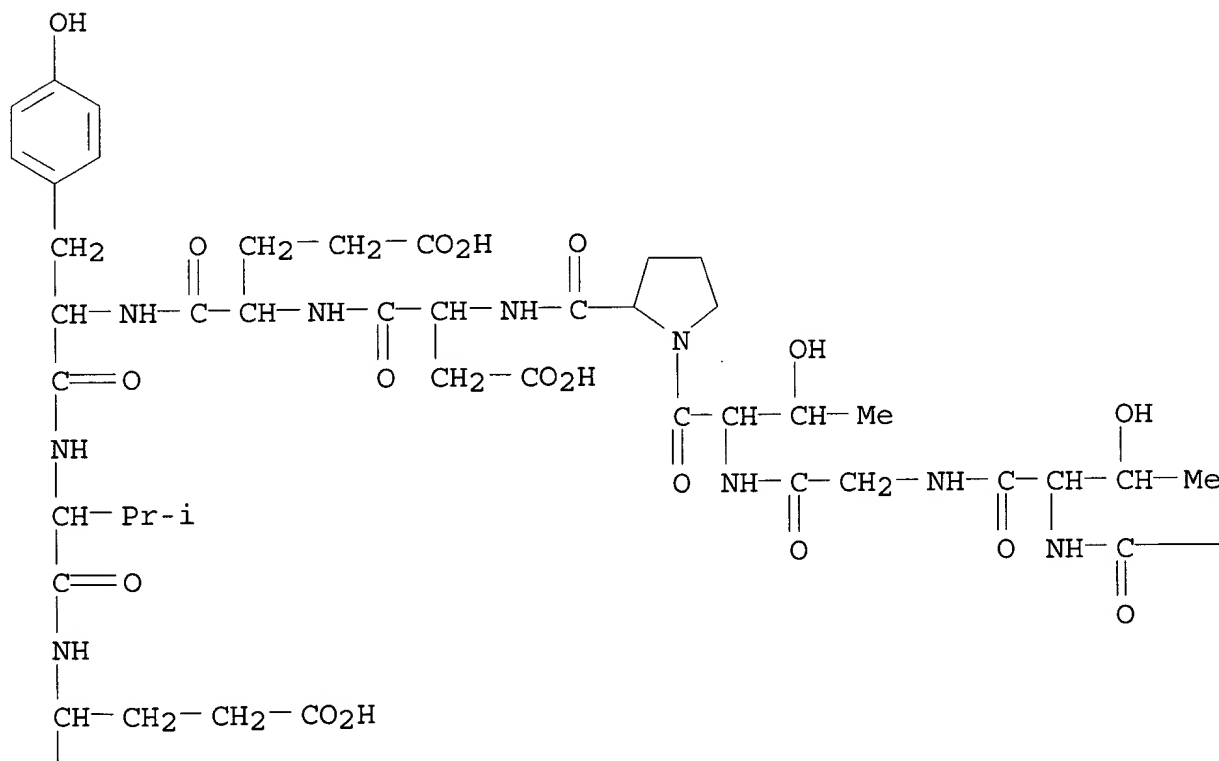
MF C146 H229 N35 045

SR	CA
----	----

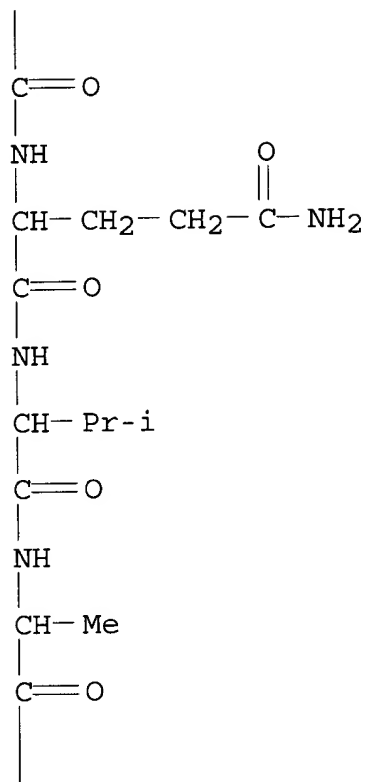
LC STN Files: CA

DES 5:ALL, L

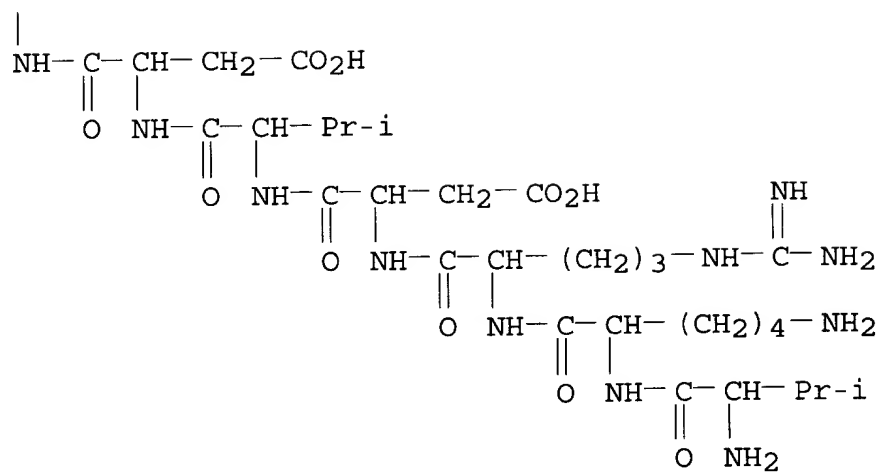
PAGE 1-A



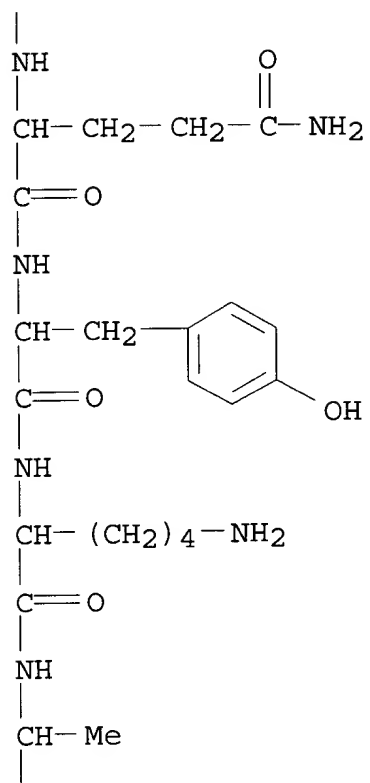
PAGE 2-A



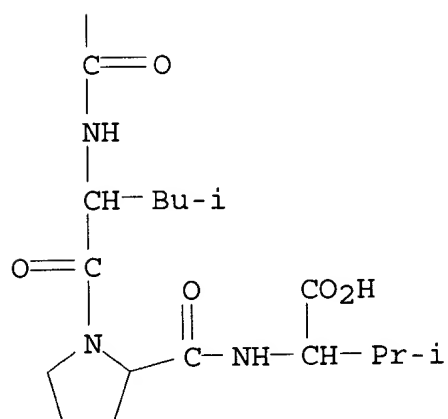
PAGE 2-B



PAGE 3-A



PAGE 4-A



1 REFERENCES IN FILE CA (1967 TO DATE)

REFERENCE 1: P 115:205920

=> fil hca

FILE 'HCA' ENTERED AT 11:23:48 ON 24 MAR 95
USE IS SUBJECT TO THE TERMS OF YOUR CUSTOMER AGREEMENT
COPYRIGHT (C) 1995 AMERICAN CHEMICAL SOCIETY (ACS)

FILE COVERS 1967 - 18 Mar 1995 (950318/ED) VOL 122 ISS 12

HCAPLUS IS NOW ONLINE!

=> d his l3

(FILE 'HCA' ENTERED AT 11:18:51 ON 24 MAR 95)
L3 5 S L2

FILE 'REGISTRY' ENTERED AT 11:19:18 ON 24 MAR 95

FILE 'REGISTRY' ENTERED AT 11:20:30 ON 24 MAR 95

FILE 'HCA' ENTERED AT 11:23:48 ON 24 MAR 95

=> d bib abs hitrn 1-5

L3 ANSWER 1 OF 5 HCA COPYRIGHT 1995 ACS
AN 121:228217 HCA
TI Potential therapeutic recombinant proteins comprised of peptides
containing recombined T cell epitopes
AU Rogers, Bruce L.; Bond, Julian F.; Craig, Sandra J.; Nault,
Anneliese K.; Segal, Debra B.; Morgenstern, Jay P.; Chen, Meei-Song;
Bizinkauskas, Christine B.; Counsell, Catherine M.; et al.
CS ImmuLogic Pharm. Corp., Waltham, MA, 02154, USA
SO Mol. Immunol. (1994), 31(13), 955-66
CODEN: MOIMD5; ISSN: 0161-5890
DT Journal
LA English
AB The complete primary structure of Fel dI has been detd. and shown to
be comprised of two sep. polypeptide chains (designated chain 1 and
chain 2). Overlapping peptides covering the entire sequence of both
chains of Fel dI have been used to map the major areas of human T
cell reactivity. The present study describes three non-contiguous T
cell reactive regions of <30 aa in length that were assembled in all
six possible configurations using PCR and recombinant DNA methods.
These six recombinant proteins comprised of defined non-contiguous T
cell epitope regions artificially combined into single polypeptide
chains have been expressed in E. coli, highly purified, and examd.
for their ability to bind to human cat-allergic IgE and for human T
cell reactivity. Several of these recombined T cell epitope-contg.
polypeptides exhibit markedly reduced IgE binding as compared to the
native Fel dI. Importantly, the human T cell reactivity to
individual T cell epitope-contg. regions is maintained even though
each was placed in an unnatural position as compared to the native
mol. In addn., T cell responses to potential junctional epitopes
were not detected. It was also demonstrated in mice that s.c.

injection of T cell epitope-contg. polypeptides inhibits the T cell response to the individual peptides upon subsequence challenge in vitro. Thus, these recombinant T cell epitope-contg. polypeptides, which harbor multiple T cell reactive regions but have significantly reduced reactivity with allergic human IgE, constitute a novel potential approach for desensitization to important allergens.

IT 136380-56-8 136380-69-3

(potential therapeutic recombinant proteins comprised of peptides contg. recombinant T cell epitopes from allergens)

L3 ANSWER 2 OF 5 HCA COPYRIGHT 1995 ACS

AN 120:75451 HCA

TI Peptides useful for inducing immune tolerance

IN Gefter, Malcolm L.; Garman, Richard D.; Greenstein, Julia L.; Kuo, Mei Chang; Briner, Thomas J.; Morville, Malcolm

PA Immunologic Pharmaceutical Corp., USA

SO PCT Int. Appl., 107 pp.

CODEN: PIXXD2

PI WO 9319178 A2 930930

DS W: AU, CA, FI, HU, JP, KP, KR, NO, NZ, PT

RW: AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE

AI WO 93-US2462 930325

PRAI US 92-857311 920325

US 92-884718 920515

US 93-6116 930115

DT Patent

LA English

AB A compn. contg. .gtoreq.1 peptide derived from a human T-cell-reactive feline protein, a protein antigen, an allergen, or an autoantigen is used for treating a disease which involves an immune response to the feline protein, protein antigen, allergen, or autoantigen. The peptides comprise a sufficient percentage of the T cell epitopes of an antigen (from Felis, Ambrosia, etc.), allergen (Der p I, Der f I, etc.) or autoantigen (insulin, myelin basic protein, etc.). Thus, s.c. administration of a combination of Lys-Arg-Asp-Val-Asp-Leu-Phe-Leu-Thr-Gly-Thr-Pro-Asp-Glu-Tyr-Val-Glu-Gln-Val-Ala-Gln-Tyr-Lys-Ala-Leu-Pro-Val and Lys-Ala-Leu-Pro-Val-Val-Leu-Glu-Asn-Ala-Arg-Ile-Leu-Lys-Asn-Cys-Val-Asp-Ala-Lys-Met-Thr-Glu-Glu-Asp-Lys-Glu (peptides derived from T-cell-reactive feline protein) induced T cell tolerance in mice.

IT 136380-56-8 136380-69-3

(T-cell tolerance induction with)

IT 136380-56-8D, Human T-cell-reactive feline protein-derived peptide mixts. 136380-69-3D, Human T-cell-reactive feline protein-derived peptide mixts.

(for T-cell tolerance induction)

IT 136796-96-8, Glycoprotein TRFP (Felis catus chain 1 isoform A protein moiety reduced) 136796-97-9, Glycoprotein TRFP (Felis catus chain 1 isoform B protein moiety reduced)
(peptide derived from, for T-cell tolerance induction)

L3 ANSWER 3 OF 5 HCA COPYRIGHT 1995 ACS
AN 119:93527 HCA
TI Recombitope peptides containing T cell epitopes and stimulating T
cell activity, for allergy therapy and diagnosis
IN Rogers, Bruce L.; Morgenstern, Jay P.; Bond, Julian F.; Garman,
Richard D.; Kuo, Mei Chang; Morville, Malcolm
PA Immulogic Pharmaceutical Corp., USA
SO PCT Int. Appl., 73 pp.
CODEN: PIXXD2
PI WO 9308280 A1 930429
DS W: AU, CA, FI, HU, JP, KR, NO
RW: AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, SE
AI WO 92-US8694 921016
PRAI US 91-777859 911016
US 91-807529 911213
DT Patent
LA English
AB Recombitope peptides, stimulating T cell activity and comprising
.gtoreq.2 T cell epitopes derived from the same or from different
protein antigens, are provided. The peptides can be derived from
protein allergens, autoantigens, or other protein antigens. Methods
of diagnosing sensitivity to an allergen or other protein antigen,
methods to treat such sensitivity, methods for designing recombiteope
peptides where the protein antigen has unknown or ill-defined T cell
epitopes, and therapeutic compns. are also disclosed. T cell
epitopic studies were done with peptides and protein chains of the
human T cell-reactive feline protein (TRFP) and immunoreactive
regions were identified. Synthetic oligonucleotides were designed
with Escherichia coli-preferred codons for PCR amplification and
expression in E. coli of recombiteope peptides from TRFP. The
peptide sequences included a 6 His residue leader sequence (for
allowing purifn. of the expressed recombiteope peptide using QIAGEN
NTA-agarose) and a thrombin cleavage site before the actual
recombiteope sequence. Recombiteope peptide arrangements were
identified which had little to no binding to IgE and which gave
responses to T cells of patients allergic to TRFP.
IT 136796-96-8, Leader A-human T cell-reactive feline protein
chain 1 (cat) 136796-97-9, Leader B-human T cell-reactive
feline protein chain 1 (cat)
(amino acid sequence of and T cell epitopes-contg. T
cell-stimulating recombiteope peptides recombinant prepn. in
relation to)
IT 136380-69-3, Human T cell-reactive feline protein chain 1
(29-55) (cat synthetic)
(recombiteope peptide contg., cat allergy detection and treatment
with)

L3 ANSWER 4 OF 5 HCA COPYRIGHT 1995 ACS
AN 118:122542 HCA
TI Amino acid sequence of Fel dI, the major allergen of the domestic
cat: protein sequence analysis and cDNA cloning